Lian DF, Liu Q. Jointly recommending library books and predicting academic performance: A mutual reinforcement perspective. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 33(4): 654–667 July 2018. DOI 10.1007/s11390-018-1847-y

# Jointly Recommending Library Books and Predicting Academic Performance: A Mutual Reinforcement Perspective

De-Fu Lian[1], *Member, CCF, ACM, IEEE*, and Qi Liu[2], *Member, CCF, IEEE*

[1] *School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731, China*

[2] *School of Computer Science and Technology, University of Science and Technology of China, Hefei 230022, China*

E-mail: {dove.ustc, liuqiah}@gmail.com

Received January 2, 2018; revised May 29, 2018.

**Abstract** The prediction of academic performance is one of the most important tasks in educational data mining, and has been widely studied in massive open online courses (MOOCs) and intelligent tutoring systems. Academic performance can be affected by factors like personality, skills, social environment, and the use of library books. However, it is still less investigated about how the use of library books can affect the academic performance of college students and even leverage book-loan history for predicting academic performance. To this end, we propose a supervised content-aware matrix factorization for mutual reinforcement of academic performance prediction and library book recommendation. This model not only addresses the sparsity challenge by explainable dimension reduction techniques, but also quantifies the importance of library books in predicting academic performance. Finally, we evaluate the proposed model on three consecutive years of book-loan history and cumulative grade point average of 13 047 undergraduate students in one university. The results show that the proposed model outperforms the competing baselines on both tasks, and that academic performance not only is predictable from the book-loan history but also improves the recommendation of library books for students.

**Keywords** book-borrowing record, educational data mining, matrix factorization, multi-task learning, student performance prediction, transfer learning

## 1 Introduction

Since course failure largely affects students' graduation, job seeking, and even future development, it becomes a great concern of higher educational management. Early prediction of academic performance may warn students against the happening of potential course failure, notify educators and administrators of in-time intervention, and thus probably prevent delivering adverse consequence to students.

The prediction of academic performance has been widely studied in intelligent tutoring systems. Based on students' interaction logs with intelligent tutoring systems, it is possible to analyze students' knowledge of skills based on student models like knowledge tracing using Hidden Markov Model[1-2] and Recurrent Neural Network[3], and like cognitive diagnostic models using deterministic inputs, noisy "And" — DINA[4] and item response theory[5-6]. It is also possible to assign skills to each question based on (non-negative) matrix factorization[7-9], singular value decomposition[10], and Q-matrix[11]. In other words, student modeling aims to find out the strength and weakness of students based on their response to questions. Massive open online courses (MOOCs) such as Coursera and Edx have become increasingly popular recently and provide students the opportunity to take online courses from prestigious universities, leading the worldwide revolution

Regular Paper

Special Section on Recommender Systems with Big Data

A preliminary version of the paper was published in the Proceedings of ICDM 2016.

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61502077 and 61672483, and the Fundamental Research Funds for the Central Universities of China under Grant No. ZYGX2016J087.

of education. As all students' learning behaviors take place on the Web, based on the recorded data of learning behaviors, it is possible to evaluate students' performance in a more objective and quantitative way. Because MOOCs are facing the low completion rates (less than 5%) of participants, one of the most important tasks in MOOCs is to reveal important factors affecting students' dropout[12-13], and develop appropriate strategies to retain students in a course. According to previous studies[14-16], behaviors about video watching, assignments attempting or quizzes taking, forums posting/viewing/replying and peer friendship, can play an important role in students' learning performance.

In secondary school or distant education, students' demographics, personality, class-attendance records, test/quiz grades and past performance history, have been leveraged for predicting academic outcomes[17-19] based on supervised learning techniques. However, these solutions are not generally applicable in the modern university, because 1) some important data, such as class-attendance records and quiz grades, is rarely digitized, and 2) some other information, such as demographics, and collected past performance history per term, is comparably static, being unable to reflect in-time change of academic performance. However, with the recent development of information technology, the computerized level in the modern university continues to increase, indicating a clear trend for the digitization of students' behaviors. This makes it possible to predict future academic performance based on these sources of information.

The usage of library books has shown significant contribution to academic success and/or student retention according to past research [20-22]. These results can be further verified by basic statistics in Fig.1(a), which indicates that students at different performance levels borrow different numbers of books from library. Such statistical analysis is too coarse to answer the questions like "which books are positively/negatively correlated with students' academic performance". If the relationship between library book usage and academic performance is analyzed at a level of books, we suffer from the sparsity challenge, since each student only borrows a small number of books from library. Therefore, there still lacks a systematic framework for mining the book-loan history to predict academic performance. To this end, we propose a supervised content-aware matrix factorization framework for the predictive analysis of academic performance based on the book-loan history and students' academic performance.

To address the sparsity challenge, this predictive framework exploits dimension reduction techniques based on content-aware matrix factorization for extracting book-loan preference of each student. The usage of content-aware matrix factorization lies in the easy way of incorporating supervised signals and the convenient way of taking book-loan data as implicit feedback. It then feeds students' preference into a regression algorithm with multi-task learning for predicting their academic performance. The reason why to equip with multi-task learning is motivated by the distinct preference of students at different schools, as shown in Fig.1(b). For example, students from computer science prefer to borrow books of the TP category. When regression coefficients are fixed, students' book-loan preference can be refined by the supervised information of academic performance. Such an alternative iteration procedure, whose complexity is in linear proportion to the size of the book-loan history,

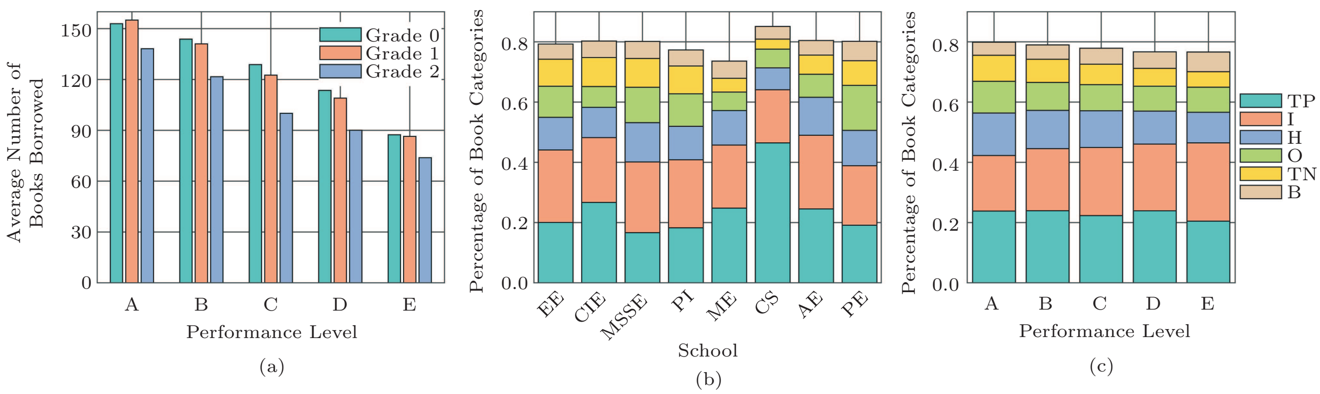



Fig.1. (a) Averaging book-loan frequency at different performance levels (A(best)∼E(worst)) and book category distribution at (b) different schools (such as Computer Science and Economic Management ) and (c) different performance levels. Category abbreviation refers to Chinese library classification. For example, TP is "Automation, Computer Engineering" and I is "literature".

continues until the convergence of students' book-loan preference. Therefore, this predictive framework, on one hand, predicts academic performance based on distinct book-borrowing preference of students at different performance levels, as exemplified in Fig.1(c), and on the other hand, promotes library book recommendation by recommending "right" books for students based on their performance levels and books' meta information, making it possible to help students improve their academic performance and to alleviate low usage rate of books in modern university library[23]. To the best of our knowledge, this is the first work of jointly modeling book borrowing preference and predicting academic performance.

Based on the latent representation of students, books, and books' meta information in supervised content-aware matrix factorization, we derive a precise prediction formula for academic performance. This formula explicitly takes the effect of similar books into account and thus explains the benefit of dimension reduction based on content-aware matrix factorization. More importantly, it becomes possible to quantify the importance of library books in predicting academic performance based on regression with uncertainty.

Finally, we evaluate the proposed model on a real-world dataset of 13 047 undergraduate students in one university, including three consecutive years of book-loan history with 676 757 records and cumulative grade point average over these three years. The experimental results indicate that the proposed algorithm outperforms the competing baselines on both tasks, and that academic performance not only is predictable from the book-loan history but also promotes the effectiveness of book recommendation.

This article is an extension of our preliminary work[24], in which the contributions are summarized as follows.

• We conducted the first systematic study for jointly recommending library books based on academic performance and books' meta information, and predicting academic performance based on the book-loan history. This makes it possible to help students improve their academic performance by recommending more personalized books for students.

• We proposed a supervised content-aware matrix factorization algorithm with multi-task learning to address the sparsity challenge of the book-loan history. And we further derive a prediction formula for academic performance, explaining the reason why dimension reduction can take effect.

In this article, we further deliver the following new contributions.

• Based on performance regression with uncertainty using book latent representation, we propose a hierarchical Bayesian model for quantifying the importance of library books in predicting academic performance.

• We perform in-depth analysis to figure out negatively and positively correlated books with academic performance for three different schools including School of Computer Science, Foreign Language, and Public and Politician Management.

## 2 Related Work

Based on the studied topics and used techniques, we present related work on academic performance prediction, inference of traits and attributes, and book and article recommendation.

### 2.1 Academic Performance Prediction

The related work with academic performance prediction comprises the following three categories. The first category involves predicting educational outcomes in high schools or colleges, such as cumulated grades or at risk of failure or dropout, based on students' demographics, past test performance and course attending history[17-18,25]. Most of them revealed that past testing performances were highly predictive of future success/failure while a set of demographic features were able to achieve reasonable high prediction accuracies as well. The second category predicts students' performance on specific problems based on interaction logs with intelligent tutoring systems or students' response to problems with concepts/skills specified by teachers, which record each problem-solving step. This involves two types of techniques. One is cognitive analysis, assigning concepts/skills to each problem, based on (non-negative) matrix factorization[8,26] or hill-climbing optimized Q-matrix methods. After mapping concepts/skills to each problem, we obtain a Q-matrix to capture this mapping. Based on longitudinal students' response to each problem, the other technique involved is student modeling, to estimate the knowledge of skills based on (dynamic) cognitive diagnosis model[4,27-28] or based on Bayesian knowledge tracing[1] and deep knowledge tracing[3]. The third category addresses the low complete rate of online courses, by early identifying students at risk of not completing online courses based on engagement activities, such as watching lectures, attempting assignments/quizzes, and posting/viewing in forums[29-30] or based on the developed/learned individual engagement taxonomy[14-16] from the engagement activities. These studies revealed several intriguing discoveries. For example, hard working and frequent questioning do not necessarily imply high learning performance, but engaging in the course forum is a significant indicator for students' learning performance.

### 2.2 Inference of Traits and Attributes

This work can also be summarized as inferring personal traits and attributes from digital records. Thus related work includes the study leveraging Facebook likes to predict different traits and attributes, such as sexual orientation, ethnicity, personality traits, intelligence, age and gender[31]. In this study, they exploited singular value decomposition for dimension reduction and fed user preference vector into linear or logistic regression for prediction. The prediction of personality can be judged more accurately by Facebook likes than that made by human[32]. In addition to Facebook likes, other digital records such as human mobility data[33], social network activities[34], website traffic data[35], and webpage browsing information[36] were also used for inferring various demographic attributes. Most of them also leveraged dimension reduction techniques for learning the low dimensional representation of user preference. For example, the factorization of user-context-knowledge tensor was proposed for inferring the users' demographics (ender, age, education background, and marital status) based location check-ins[33]. Singular value decomposition was applied on user-webpage click matrix for extracting user preference for webpages, and then fed together with pages' content-based and category-based features into support vector machine regression model for predicting gender and age. Although many digital records have been utilized, the book-borrowing records as very informative and fundamental student behaviors in a campus have by and large been overlooked for academic performance prediction. Moreover, most previous work predicted traits and attributes directly using the extracted preference from dimension reduction, and thus the prediction of traits and attributes is independent to the extraction of user preference, being different from our supervised dimension reduction techniques.

### 2.3 Book and Article Recommendation

Book recommendation is one of tasks in this paper and has been well studied using the book-loan, book-reviewing or book-purchase history in the field of recommendation system. For example, a Naive Bayesian classifier was designed for the recommendation of books based on content filtering[37] using the book-reviewing history. Constructing a two-layer graph which consists of a book-book and user-user relation graph and is connected by user-book relationship, a graph search based hybrid recommendation approach was proposed for book recommendation[38] using the book-purchase history. Recently, a challenge about book recommendation based on the book-reviewing history has been launched. In this challenge, most participants used hybrid approaches[39], such as stack regression and rank aggression, incorporating book features into matrix to factorize and conclude the significant benefit of book contents. Similar to book recommendation, book contents also play important role in better-studied article recommendation, so that the superior algorithm of article recommendation can be potentially used for book recommendation. Taking some superior algorithms of article recommendation for example, in [40], traditional collaborative filtering was integrated with probabilistic topic modeling based on additive models; to improve the representation power of article contents, stacked denoising auto-encoder (SDAE) was leveraged for reducing the dimension of article representation and integrated with traditional collaborative filtering[41] in a joint optimization framework. However, not only book contents, but also their information such as categories and authors of books could be useful. In order to take all of these information into account, we exploit a more general framework for content-aware collaborative filtering[42-46] or for tensor factorization[47-49]. Moreover, all these algorithms do not consider students' academic performance for improving book recommendation, and thus a supervised content-aware matrix factorization is proposed in this paper.

## 3 Overview and Preliminary

In this paper, academic performance is predicted based on students' book-loan history. Since each student only borrows a small number of books from library, considering each book as feature index for academic performance prediction suffers from the data scarcity problem. Instead, a dimension reduction technique should be applied to extract students' borrowing preference. These learned preferences are then considered as features and fed into regression techniques for academic performance prediction. However, without book reviews after returning, students' negative preference for borrowed books cannot be reflected. Thus, the book-loan history is a kind of implicit feedback, whose loan frequency determines the confidence of positive preference. In this case, weighted matrix factorization becomes an optimal choice for dimension reduction, due to the superiority in implicit feedback[50]. Below, we first make a brief introduction to them.

### 3.1 Weighted Matrix Factorization

The proposed algorithm operates on a student-book loan matrix $\boldsymbol{R} \in \{0,1\}^{M\times N}$, including $M$ students and $N$ books, based on a confidence matrix $\boldsymbol{W} \in \mathbb{R}^{M\times N}$. Each entry $r_{i,j}$ in matrix $\boldsymbol{R}$ indicates whether a student $i$ has borrowed a book $j$ or not, and each $w_{i,j}$ in matrix $\boldsymbol{W}$ indicates his/her preference confidence for this book. The confidence of all non-borrowed books is assigned to 1 while the confidence of borrowed ones is assigned to a value being significantly larger than 1 and monotonic with the loan frequency. In this setting, weighted matrix factorization achieves dimension reduction by optimizing the following objective function:

$$\mathcal{L} = \sum_{i,j} w_{i,j}(r_{i,j} - \tilde{\boldsymbol{p}}_i'\tilde{\boldsymbol{q}}_j)^2 + \alpha\Big(\sum_i \|\tilde{\boldsymbol{p}}_i\|^2 + \sum_j \|\tilde{\boldsymbol{q}}_j\|^2\Big), \quad (1)$$

where $\tilde{\boldsymbol{p}}_i \in \mathbb{R}^K$ is a latent vector of a user $i$ and $\tilde{\boldsymbol{q}}_j \in \mathbb{R}^K$ is a latent vector of a book $j$ so that both students and books are mapped into a joint latent space, where the dot product between their latent vectors indicates students' borrowing preference for books.

### 3.2 Content-Aware Weighted Matrix Factorization

When students are accompanied by profiles, and books are provided content information, such as categories and prefaces, content-aware weighted matrix factorization[42,46] should be suggested. This algorithm first represents the profile of each student by a feature vector $\boldsymbol{x} \in \mathbb{R}^F$ of $F$ features, represents the content of each book by a feature vector $\boldsymbol{y} \in \mathbb{R}^L$ of $L$ features, respectively, and then maps them into the same joint latent space as generated by weighted matrix factorization by multiplying feature latent matrices $\boldsymbol{U} \in \mathbb{R}^{F\times K}$ and $\boldsymbol{V} \in \mathbb{R}^{L\times K}$. Therefore, they can be directly added into latent factors of students and books, i.e., $\boldsymbol{p}_i = \tilde{\boldsymbol{p}}_i + \boldsymbol{U}\boldsymbol{x}_i$ and $\boldsymbol{q}_j = \tilde{\boldsymbol{q}}_j + \boldsymbol{V}\boldsymbol{y}_j$. After substituting them into the objecting function (1) and regularizing $\boldsymbol{U}$ and $\boldsymbol{V}$ by Frobenius norms, content-aware weighted matrix factorization optimizes the objective function as follows:

$$\mathcal{L}_{DR} = \sum_{i,j} w_{i,j}(r_{i,j} - \boldsymbol{p}_i'\boldsymbol{q}_j)^2 + \alpha\sum_i \|\boldsymbol{p}_i - \boldsymbol{U}'\boldsymbol{x}_i\|^2 + \alpha\sum_j \|\boldsymbol{q}_j - \boldsymbol{V}'\boldsymbol{y}_j\|^2 + \beta(\|\boldsymbol{U}\|_F^2 + \|\boldsymbol{V}\|_F^2).$$

Since this objective function is quadratic with respect to each variable when others are fixed, it can be optimized by alternating least square algorithm. Their updating formula will be elaborated below due to its common with the proposed algorithm.

## 4 Supervised Content-Aware Weighted Matrix Factorization with Multi-Task Learning

Based on the content-aware weighted matrix factorization, we represent each user by a latent factor, which not only captures students' borrowing preference but also absorbs his/her profile information. This user representation is both able to predict a student's academic performance by considering it as a feature vector in a supervised learning model, and able to recommend books based on its dot product with book latent factors. However, such a paradigm neither makes sure the extracted features by dimension reduction are optimal for academic performance prediction, nor renders book recommendation benefit from students' performance information. Therefore, we propose a Supervised Content-aware Weighted Matrix Factorization with Multi-Task Learning (SCWMF-MTL) for jointly predicting academic performance and recommending library books, that is, iteratively learning student preference from borrowing history based on content-aware matrix factorization and the supervision of student academic performance, and updating the parameters of the prediction model of academic performance, until the convergence of user preference factors.

### 4.1 Loss Function

Before presenting this model, we first set up the task of predicting academic performance. Provided cumulative grade point average (CGPA) of each student, we first subtract it by the mean CGPA of all students at the same schools, since this group of students take extremely similar courses. Although the difficulty of courses is varied from school to school, and their lecturers' teaching skills may be different from one another, they can be eliminated by this preprocessing, making students' CGPA be comparable with one another. Setting the deduced CGPA $z_i$ of each student $i$ as regressand, and considering his/her latent factor $\boldsymbol{p}_i$ as regressor, we can apply regression techniques for academic performance prediction.

However, the same aspects of latent factors play a different role in predicting academic performance among different schools, just as illustrated in Fig.1(c). For example, students at the School of Humanities and Social Science can benefit from reading novels while students at the School of Computer Science may not. Therefore, regression weights should be varied from school to school, but learning school-specified weights suffers from the over-fitting problem. To this end, we apply a multi-task learning algorithm for academic performance regression, so that school-specified weight of different schools can be shared to some extent with one another.

Based on this above setting, assuming there are $S$ schools in the university, we can formulate the loss function of SCWMF-MTL as follows:

$$\mathcal{L} = \sum_i (z_i - \boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{p}_i)^2 + \lambda_D \mathcal{L}_{DR} + \lambda_M \mathrm{tr}(\boldsymbol{G}'\boldsymbol{L}\boldsymbol{G}) + \lambda_R \mathrm{tr}(\boldsymbol{G}'\boldsymbol{G}),$$

where each row of $\boldsymbol{G} \in \mathbb{R}^{S\times K}$ corresponds to regression coefficients of the corresponding schools and $\boldsymbol{e}_i = (e_{i,1}, \cdots, e_{i,S})$ is a school-selection vector subject to $e_{i,s} = 1$ if a student $i$ is at a school $s$, and $e_{i,s'} = 0$, otherwise. $\boldsymbol{L} = \boldsymbol{I}_S - \frac{1}{S}\boldsymbol{1}_S\boldsymbol{1}_S'$ is a centered matrix, and thus $\mathrm{tr}(\boldsymbol{G}'\boldsymbol{L}\boldsymbol{G})$ measures the row-based variance of matrix $\boldsymbol{G}$. In other words, it makes sure that each row of matrix $\boldsymbol{G}$ is close to their mean. $\mathrm{tr}(\boldsymbol{G}'\boldsymbol{G}) = \|\boldsymbol{G}\|_F^2$ is a regularization term, avoiding the over-fitting and being controlled by $\lambda_R$.

### 4.2 Optimization

According to the analysis to this objective function, it is quadratic with respect to each variable of $\{\boldsymbol{p}_i, \boldsymbol{q}_j, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{g}_s\}$. Therefore, given the other variables fixed, we can get an analytic solution for each variable.

In particular, setting the gradient of $\mathcal{L}$ with respect to $\boldsymbol{p}_i$ to zero, we can get

$$\boldsymbol{p}_i = (\frac{\boldsymbol{g}_s\boldsymbol{g}_s'}{\lambda_D} + \boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{Q} + \alpha\boldsymbol{I}_K)^{-1} \times (\boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{r}_i + \alpha\boldsymbol{U}'\boldsymbol{x}_i + \frac{z_i\boldsymbol{g}_s}{\lambda_D}),$$

where $\boldsymbol{W}^i = \mathrm{diag}(w_{i,1}, \cdots, w_{i,N})$ and we assume student $i$ is at school $s$. Due to the setting of the dense weight matrix, this updating formula can be efficient since $\boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{Q} = \boldsymbol{Q}'(\boldsymbol{W}^i - \boldsymbol{I}_N)\boldsymbol{Q} + \boldsymbol{Q}'\boldsymbol{Q}$ and $\boldsymbol{W}^i - \boldsymbol{I}_N$ is a sparse matrix, whose number of entries equals $\|\boldsymbol{r}_i\|_0$. $\boldsymbol{Q}'\boldsymbol{Q}$ is independent to users and can thus be precomputed before each user's update. Therefore, the complexity of updating latent factor of user $i$ is $\mathcal{O}(\|\boldsymbol{r}_i\|_0 K^2 + K^3)$, and thus the total complexity is $\mathcal{O}(\|\boldsymbol{R}\|_0 K^2 + MK^3)$.

Setting the gradient of $\mathcal{L}$ with respect to $\boldsymbol{g}_s$ to zero, we can obtain the updating formula for $\boldsymbol{g}_s$:

$$\boldsymbol{g}_s = (\boldsymbol{P}'\boldsymbol{E}^s\boldsymbol{P} + \lambda\boldsymbol{I}_K)^{-1}(\boldsymbol{P}'\boldsymbol{E}^s\boldsymbol{z} + \frac{\lambda_M}{S}(\boldsymbol{G}'\boldsymbol{1}_S - \boldsymbol{g}_s)),$$

where $\lambda = (\lambda_M\frac{S-1}{S} + \lambda_R)$ and $\boldsymbol{E}^s = \mathrm{diag}(e_{1,s}, \cdots, e_{M,s})$. Due to the existence of the regularization term, $\mathrm{tr}(\boldsymbol{G}'\boldsymbol{L}\boldsymbol{G})$, the parameter $\boldsymbol{g}_s$ of the prediction model for the school $s$ is affected by other schools, and can play important role in alleviating the over-fitting problem resulting from data partition by schools. Following similar analysis, the updating complexity for parameters in the prediction model of all schools is $\mathcal{O}(SK^3 + MK^2)$, dominated by the inversion of the $K \times K$ matrix.

Setting the gradient of $\mathcal{L}$ with respect to $\boldsymbol{q}_j$ to zero, we can get the updating formula for book latent factor:

$$\boldsymbol{q}_j = (\boldsymbol{P}'\boldsymbol{W}^j\boldsymbol{P} + \alpha\boldsymbol{I}_K)^{-1}(\boldsymbol{P}'\boldsymbol{W}^j\boldsymbol{r}_j + \alpha\boldsymbol{V}'\boldsymbol{y}_j),$$

where $\boldsymbol{W}^j = \mathrm{diag}(w_{1,j}, \cdots, w_{M,j})$. Similar to the update of $\boldsymbol{p}_i$, this updating formula could be efficiently implemented and the overall complexity of update is $\mathcal{O}(\|\boldsymbol{R}\|_0 K^2 + NK^3)$.

Setting the gradient of $\mathcal{L}$ with respect to $\boldsymbol{U}$ and $\boldsymbol{V}$ to zero, we can get the analytic solution of $\boldsymbol{U}$ and $\boldsymbol{V}$, that is,

$$\boldsymbol{U} = (\boldsymbol{X}'\boldsymbol{X} + \frac{\beta}{\alpha}\boldsymbol{I}_F)^{-1}\boldsymbol{X}'\boldsymbol{P},$$
$$\boldsymbol{V} = (\boldsymbol{Y}'\boldsymbol{Y} + \frac{\beta}{\alpha}\boldsymbol{I}_L)^{-1}\boldsymbol{Y}'\boldsymbol{Q}.$$

According to the complexity of matrix multiplication and inversion of matrix, the overall complexity of updating them is $\mathcal{O}(NF^2 + F^3 + NFK + ML^2 + L^3 + MLK)$. When the number of features is small, this updating formula is efficient. When the number of features is large, we need resort to conjugate gradient descent, whose complexity is $\mathcal{O}(\|\boldsymbol{X}\|_0 + \|\boldsymbol{Y}\|_0)K\#iter$ according to [42], where $\#iter$ is the number of iterations of conjugate gradient descent to reach a given threshold of approximation error.

Given these updating formulas, we then perform learning these parameters by alternating optimization, that is, taking turns to update each variable, until the convergent of $\mathcal{L}$. In addition, the latent factors of each user and each book are updated independently and can be achieved in a parallel way, but the regression coefficient of each school depends on each other so that the order of updating should be randomized.

*Complexity Analysis*. Assume conjugate gradient descent is applied for getting the solution of $\boldsymbol{U}$ and $\boldsymbol{V}$, the complexity of updating $\{\boldsymbol{p}_i, \boldsymbol{q}_j, \boldsymbol{U}, \boldsymbol{V}\}$ in one round is $\mathcal{O}\big((\|\boldsymbol{X}\|_0 + \|\boldsymbol{Y}\|_0)K\#iter + \|\boldsymbol{R}\|_0 K^2\big)$, since the updating cost for $\boldsymbol{g}_s$ of all schools in one round is usually significantly smaller than the former part. Therefore, the optimization algorithm is scalable with the size of the book-loan history and the size of the user-feature matrix and the item-feature matrix.

### 4.3 Explainable Academic Performance Prediction

After learning latent factors of students and books as well as their features, and learning the regression coefficients, we next present how to predict academic performance based on these parameters. For the sake of reasonable evaluation, students in the training dataset are assumed disjointed with students in the testing dataset, and thus latent factors of training users are useless in predicting the academic performance of testing students. The latent factors of each testing user are required first to learn and then fed into the linear predicting function. Therefore, the final formula of academic performance prediction for a testing user $i$ is represented as

$$\tilde{z}_i = \boldsymbol{e}_i'\boldsymbol{G}(\boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{Q} + \alpha\boldsymbol{I}_K)^{-1}(\boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{r}_i + \alpha\boldsymbol{U}'\boldsymbol{x}_i), \quad (2)$$

where $\alpha$ and $\boldsymbol{W}^i$ follow the same setting as the training phase. Delving into (2), we observe the prediction score involves the addition of two parts, where one part is related to the book-loan history and the other part relies on the features of students. Applying the Woodbury matrix identity for inverting the matrix $\boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{Q}+\alpha\boldsymbol{I}_K$, we can rewrite the predictive function as follows:

$$\begin{aligned}\tilde{z}_i = {} & \boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{r}_i/\alpha + \boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{U}'\boldsymbol{x}_i - \\ & \boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{Q}'((\alpha\boldsymbol{W}^{-i} + \boldsymbol{Q}\boldsymbol{Q}')^{-1}\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{W}^i\boldsymbol{r}_i/\alpha + \boldsymbol{U}'\boldsymbol{x}_i)),\end{aligned}$$

where $\boldsymbol{W}^{-i} \triangleq (\boldsymbol{W}^i)^{-1}$ for simplifying notations. This prediction involves a linear function of four different types of features. The first type of features is the books borrowed by students, weighted by $\boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{Q}'$. The second type depends on the features of students, weighted by $\boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{U}'$. The third type is the similar books to what they borrow, also weighted by $\boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{Q}'$. The similarity between books is expressed by $(\alpha\boldsymbol{W}^{-i} + \boldsymbol{Q}\boldsymbol{Q}')^{-1}\boldsymbol{Q}\boldsymbol{Q}'$. And the final type is the books preferred by a student population having the same features as student $i$, also weighted by $\boldsymbol{e}_i'\boldsymbol{G}\boldsymbol{Q}'$. From these four types of features, it is obvious to understand the benefit of dimension reduction, that is, to consider not only the books borrowed by students themselves but also the similar books to what they borrow.

### 4.4 Book Predictive Power Modeling

Based on the above model, we can obtain the regression coefficients for student latent factors in the task of academic performance prediction. However, since we do not directly use books for prediction, it is unclear about how to measure the predictive powers of books. This is important in performance diagnosis, i.e., better understanding which books could potentially improve/deteriorate academic performance. Actually, an analogy problem has been raised and discussed in [31], where the authors measured the "importance" of each book in our scenario by averaging the academic performance of all book-borrowers. For example, the book $b_3$ in Fig.2 is borrowed by four students $\{u_1, u_2, u_3, u_4\}$, and the "importance" of this book is $\frac{1}{4}\sum_{i=1}^{4} z_i$. Therefore, higher absolute averaged academic performance indicates higher importance of books in academic performance. However, only a small portion of books can evaluate their predictive power in this method, because each book is assumed to be borrowed by at least 100 students in order to robustly estimate the averaging academic performance. According to the statistics to the loan-history, the number of books borrowed by at least 100, 70 and 50 students is 10, 103 and 460, respectively. In other words, if we follow previous approaches, only 10 books can show their importance in predicting academic performance.

In order to assess the importance of any book, even if only borrowed by a few students, we propose a book predictive power modeling approach based on the extracted book latent representation from supervised content-aware matrix factorization. This approach will establish a linear mapping function $f(\boldsymbol{q}) = \boldsymbol{h}'\boldsymbol{q}$ from book latent representation $\boldsymbol{q}$ to academic performance of each book. We assume that each student's academic performance can be affected by the books they borrowed, and the affecting level depends on books themselves. Denoting $\mathbb{U}_j$ as the students borrowing the book $j$, we need to optimize the following hierarchical Bayesian model for getting $\boldsymbol{h}$,

$$\prod_{j=1}^{N}\prod_{i\in\mathbb{U}_j}\mathcal{N}(z_i|\boldsymbol{h}'\boldsymbol{q}_j,\lambda_j^{-1})\mathrm{Ga}(\lambda_j|a_0,b_0)\mathcal{N}(\boldsymbol{h}|0,\sigma^2),$$

where $\mathrm{Ga}(\lambda_j|a_0,b_0)$ means $\lambda_j$ is subject to Gamma distribution. If we reorganize this objective by users instead of books, the performance $z_i$ of student $i$ is subject to $\mathcal{N}(z_i|\mu_i,\sigma_i^2)$, where $\mu_i=\frac{\sum_{j\in\mathbb{B}_i}\lambda_j\boldsymbol{h}'\boldsymbol{q}_j}{\sum_{j\in\mathbb{B}_i}\lambda_j}$ and $1/\sigma_i^2=\sum_{j\in\mathbb{B}_i}\lambda_j$. In other words, the performance of each student is a weighted performance sum of his/her borrowed books.

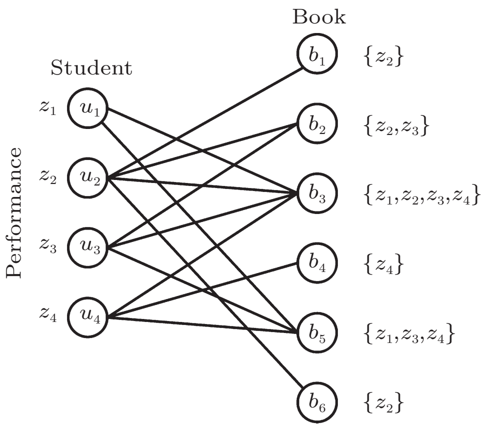



Fig.2. Book performance illustration. Each book is borrowed by several students and the performance of each book is related to a set of borrowers' academic performance.

This objective function is required to maximize with respect to $\boldsymbol{h}$ and $\lambda_i$, which can be achieved by iterated conditional modes algorithm[51], iteratively maximizing the probability of each variable conditioned on the rest. In particular, given $\lambda_i$ of all books fixed, $\boldsymbol{h}$ can be computed as follows:

$$\boldsymbol{h}=(\sum_j\boldsymbol{q}_j\boldsymbol{q}_j'\lambda_i|\mathbb{U}_j|+\sigma^{-2}\boldsymbol{I}_k)^{-1}\sum_j\boldsymbol{q}_j\lambda_j|\mathbb{U}_j|\bar{z}_{(j)},$$

where $\bar{z}_{(j)}=\frac{1}{\mathbb{U}_j}\sum_{i\in\mathbb{U}_j}z_i$ is borrowers' averaging performance of book $j$. This equation shows that books with larger affecting level ($\lambda_j$) will take larger effect in computing $\boldsymbol{h}$.

When $\boldsymbol{h}$ is fixed, we can get the updating formula for $\lambda_j$ of each book:

$$\lambda_j=\frac{\frac{|\mathbb{U}_j|}{2}+a_0-1}{b_0+\frac{1}{2}\sum_{i\in\mathbb{U}_j}(z_i-\boldsymbol{h}'\boldsymbol{q}_j)^2}.$$

Plugging $\bar{z}_{(j)}$ into the denominator, we can decompose it into two parts $\sum_{i\in\mathbb{U}_j}(z_i-\boldsymbol{h}'\boldsymbol{q}_j)^2=\sum_{i\in\mathbb{U}_j}(z_i-\bar{z}_{(j)})^2+|\mathbb{U}_j|(\bar{z}_{(j)}-\boldsymbol{h}'\boldsymbol{q}_j)^2$. Assume $a_0=1$ and $b=0$, we can get a new formula for $\lambda_j$,

$$\lambda_j^{-1}=\frac{1}{|\mathbb{U}_j|}\sum_{i\in\mathbb{U}_j}(z_i-\bar{z}_{(j)})^2+(\bar{z}_{(j)}-\boldsymbol{h}'\boldsymbol{q}_j)^2.$$

Therefore, the affecting level of each book is determined by the performance variance and bias of approximation. The larger variance and bias will incur a small affecting level so that books with larger variance and approximation bias take smaller effect in computing $\boldsymbol{h}$.

Until the convergence of iteratively updating $\lambda_j$ and $\boldsymbol{h}$, we can predict the performance based on $f(\boldsymbol{q}_j)=\boldsymbol{h}'\boldsymbol{q}_j$ for each book $j$. Taking all books borrowed by students at the same school and ranking them by $f(\boldsymbol{q}_j)$, we can get the most important books for academic performance prediction.

## 5 Experiment

The evaluation is conducted on a dataset with 16 704 undergraduate students of 19 schools spanning three consecutive grades (denoted as G0, G1 and G2). For each student, this dataset includes his/her first three years of book-loan history and cumulative grade point averages over the first three years. Each book in the loan history contains a category in Chinese library classification①. In order to learn students' stable borrowing preference, we filter out books which have only been borrowed by less than two students, and filter out students who have borrowed less than five books and who are at new established schools. The preprocessed dataset includes 13 047 students from 14 schools. Table 1 lists the statistics of this dataset, Fig.3(a) shows the distribution of book categories, and Fig.3(b) shows the distribution of student number across schools.

**Table 1.** Statistics of the Dataset

| Name | #Students | #Books | #Records | #Books per Student |
|---|---|---|---|---|
| G0 | 4 335 | 71 122 | 242 376 | 55.9 |
| G1 | 4 434 | 72 591 | 239 869 | 54.1 |
| G2 | 4 278 | 65 183 | 194 512 | 45.5 |

Note: #: number of.

Based on this dataset, we will evaluate both academic performance prediction and library book recommendation. For the former part, we consider the following two configurations. The first one is to train the proposed model on one grade of dataset (denoted as G$i$), and to test it on the dataset of the subsequent

①https://en.wikipedia.org/wiki/Chinese_Library_Classification, May 2018.

grade (G$j$), subject to $i < j$. Denoted as G$i$→G$j$, this setting will include three cases, i.e., G0→G1, G1→G2, G0→G2. The second one is to split students from all three grades into five folds and to perform five-fold cross-validation. For the latter evaluation of book recommendation, we only exploit five-fold cross-validation, since the other case corresponds to the cold-start problem, beyond the scope of this paper. More specifically, the book-loan history of each user is split into five folds and aggregated with the same fold of the book-loan history of other users.

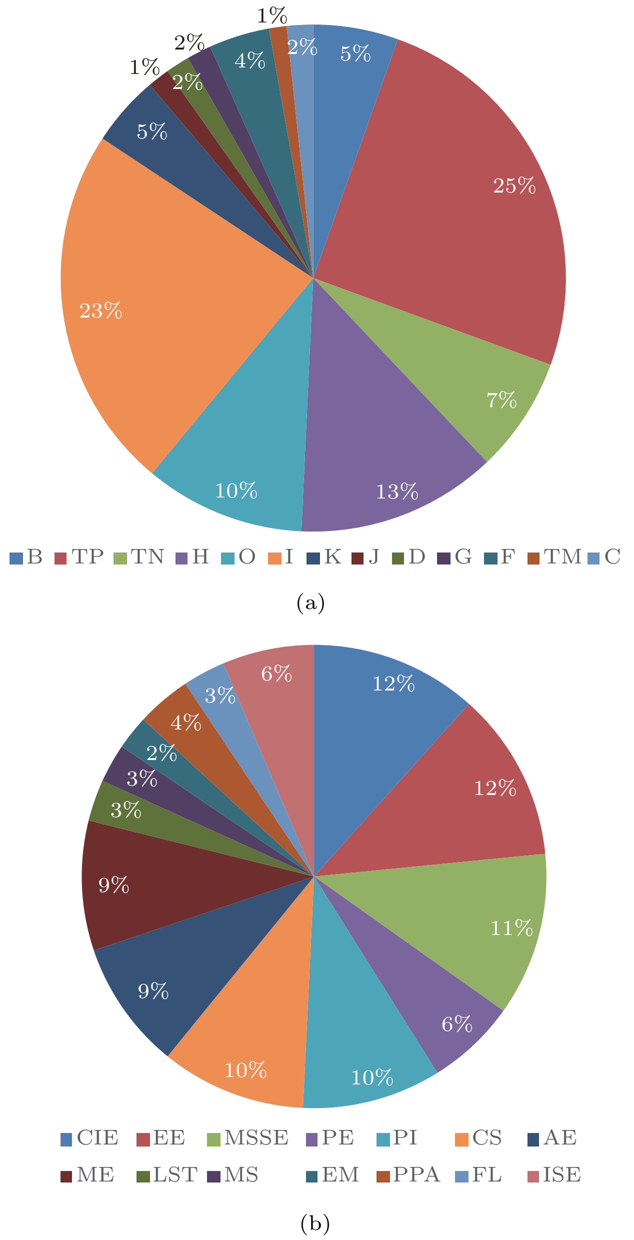



Fig.3. (a) Distribution of the top 13 categories of books borrowed by students. (b) Distribution of student number at the 14 largest schools.

### 5.1 Metric

For academic performance prediction, we quantify the model performance by measuring the consistence between the predicted order of students within the same school and the given order of students by academic performance, and averaging them over all schools. In this paper, we only consider the pairwise comparison and measure the ranking consistence within the school $s$ measured by accuracy (abbr. Acc($s$)),

$$\mathrm{Acc}(s) = \frac{\sum_{i,j\in\mathbb{U}_s} I_{((z_i - z_j)(\tilde{z}_i - \tilde{z}_j) > 0)}}{\frac{1}{2}|\mathbb{U}_s|(|\mathbb{U}_s| - 1)},$$

where $\mathbb{U}_s$ denotes the set of all students at the school $s$ and $\tilde{z}_i$ denotes the predicted score. This metric indicates the proportion of concordant pairs to all possible pairs and is strongly correlated with Kendall rank correlation coefficient, i.e., $\tau(s) = 2 \times \mathrm{Acc}(s) - 1$. A completely random guess would give 0.5 accuracy. The final predicted accuracy will be obtained by averaging Acc($s$) over all schools.

For library book recommendation, we exploit the widely-used metrics, precision and recall, at a cut-off position $k$, denoted as $prec@k$ and $recall@k$,

$$prec@k = \sum_{u=1}^{M} \frac{|\mathbb{S}_u(k) \cap \mathbb{V}_u|}{M \times k},$$

$$recall@k = \sum_{u=1}^{M} \frac{|\mathbb{S}_u(k) \cap \mathbb{V}_u|}{M \times |\mathbb{V}_u|},$$

where $\mathbb{S}_u(k)$ is the collection of top $k$ recommended books for a student $u$, and $\mathbb{V}_u$ is the set of his/her borrowed books.

### 5.2 Experimental Results

#### 5.2.1 *Academic Performance Prediction*

We will compare the proposed algorithm, i.e., SCWMF_MTL, with the following three baselines. The first one is Least_MTL, where borrowing frequency of book categories is considered as features and fed into multi-task linear regression. Its main difference from the proposed method is that the features are manually designed. The second is WMF_MTL, which first applies matrix factorization on the student-book loan matrix for learning students' borrowing preference, and then feeds them into multi-task linear regression models. Note that the factorization of training student-book loan matrix is independent to that of testing student-book loan matrix. The third one is SWMF, which does not make use of multi-task learning framework for learning parameters, compared with the proposed models. The comparison results are shown in Table 2.

From this table, we first observe that Least_MTL is not so good as SWMF and SCWMF_MTL. This shows the blindness of hand-designed features and the advantage of matrix factorization for feature extraction. However, WMF_MTL performs the worst among all studied algorithms. Therefore, the features extracting by matrix factorization just imply students' borrowing preference, but cannot reflect the difference of such preferences among students at various performance levels. By means of collaborative academic performance prediction and library book recommendation, we can extract more effective features for academic performance prediction. Finally, SCWMF_MTL outperforms SWMF, indicating the benefit of multi-task learning and confirming the difference of students' borrowing preference at different schools.

**Table 2.** Comparison with Baselines

| Accuracy | G0→G1 | G0→G2 | G1→G2 | 5-CV |
|---|---|---|---|---|
| Least_MTL | 0.566 3 | 0.569 6 | 0.571 5 | 0.578 1 |
| WMF_MTL | 0.514 2 | 0.516 1 | 0.524 9 | 0.520 0 |
| SWMF | 0.613 8 | 0.623 4 | 0.623 0 | 0.633 5 |
| SCWMF_MTL | 0.627 9 | 0.633 1 | 0.635 2 | 0.643 8 |

Note: 5-CV: 5-fold cross-validation.

To understand the benefit of collaborative learning and multi-task learning, we perform the sensitivity analysis of two important parameters, $\lambda_D$ and $\lambda_M$, where the former one indicates the trade-off between matrix factorization and multi-task learning, and the latter one implies the extent of similarity of regression coefficients among different schools. As shown in Fig.4(a), with the increase of $\lambda_D$, the performance of the proposed model first improves, and then deteriorates slightly before being stable since this collaborative process is dominated by matrix factorization. As shown in Fig.4(b), the varying trend of accuracy with the increase of $\lambda_M$ explicitly shows its optimal value and thus illustrates the effect of multi-task learning once again.

### 5.2.2 Impact of Amount of Data Available

The results presented so far rely on feeding all training loan history to learn the representation of books and regression coefficients. Therefore, it is unclear about how prediction accuracy changes with the varying number of observed loan records. Therefore, under three evaluation schemes, G0→G1, G0→G2 and G1→G2, we run the predictive model based on different percentages (ranging from 10% to 100%) of randomly selected training subsets of loan records. The results presented in Fig.5 show that only knowing 10% of training data can result in over 59% prediction accuracy in all three evaluation schemes. Knowing more loan records improves the prediction accuracy with diminishing effect from each additional portion of loan records.

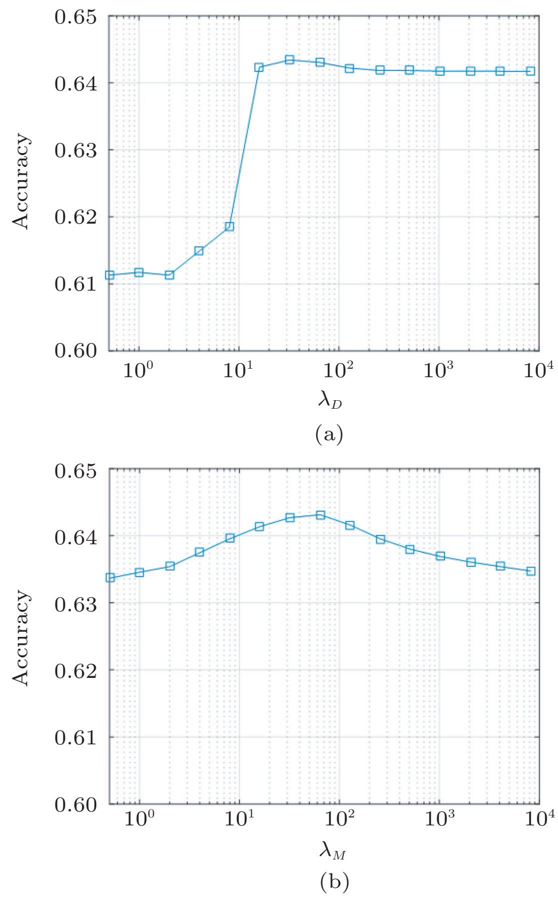



Fig.4. Sensitivity analysis of (a) $\lambda_D$ and (b) $\lambda_M$.

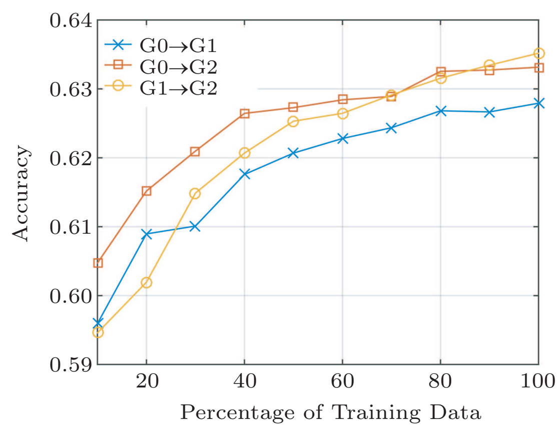



Fig.5. Impact of data availability.

*5.2.3 Book Predictive Power*

According to the definition of book importance $f(\boldsymbol{q}_j)$, we can select books highly correlated with academic performance. Based on the sign of $\boldsymbol{q}_j' \boldsymbol{g}_s$, we organize them into two groups, representing the positively correlated books and negatively correlated books, respectively. Due to the distinct loan preference of different schools, we give the results of three schools, that is, computer science (CS) in Table 3, political and public management (PPM) in Table 4, and foreign language (FL) in Table 5. For CS undergraduate students, positively correlated books include advanced materials such

**Table 3**. Top 24 Important Books Correlated with Academic Performance Based on $f(\boldsymbol{q}_j)$ of the School of Computer Science

| Order | Top 12 Positively Correlated Books | | Top 12 Negatively Correlated Books | |
|---|---|---|---|---|
| | Name of Important Books | $f(\boldsymbol{q}_j)$ | Name of Important Books | $f(\boldsymbol{q}_j)$ |
| 1 | Artificial Intelligence | 10.67 | Practical Java Training | −5.93 |
| 2 | Cryptography Theory and Practice | 10.35 | JavaScript Bible | −5.25 |
| 3 | Information Security and Privacy | 10.22 | PHP Job Collection | −4.94 |
| 4 | Engineering a Compiler | 9.59 | Mastering Spring | −4.06 |
| 5 | C++ Technology in Qt | 8.90 | Shellcoder Programming | −3.47 |
| 6 | Math Foundation for Information Security | 8.69 | Tutorial of Computer Rank Examination | −3.36 |
| 7 | Techniques & Application of Data Mining | 8.59 | Computer Game Programming | −3.10 |
| 8 | Programming Collective Intelligence | 8.47 | Freud's Psychology Philosophy | −3.04 |
| 9 | Formal Language and Automata Theory | 8.36 | Computer English | −2.44 |
| 10 | Solving ICPC Examples | 7.83 | Hacker | −2.31 |
| 11 | Operation System | 7.72 | Introduction to Node.js | −2.02 |
| 12 | Computer Architectures | 7.52 | Attack and Defense of Information Systems | −1.98 |

**Table 4**. Top 24 Important Books Correlated with Academic Performance Based on $f(\boldsymbol{q}_j)$ of the School of Political and Public Management

| Order | Top 12 Positively Correlated Books | | Top 12 Negatively Correlated Books | |
|---|---|---|---|---|
| | Name of Important Books | $f(\boldsymbol{q}_j)$ | Name of Important Books | $f(\boldsymbol{q}_j)$ |
| 1 | The Risk of Society | 9.87 | Electronic Commerce | −2.250 |
| 2 | Essentials of Western Public Admin. Theory | 9.40 | Interpretation of Hegel's Philosophy | −2.210 |
| 3 | Guiding Economic Law | 8.43 | Great American Trials | −1.960 |
| 4 | Introduction to Public Management | 8.08 | Introduction to Psychology | −1.280 |
| 5 | Global Civil Society | 7.59 | The Prince | −1.180 |
| 6 | New Institutional Economics | 7.15 | Apartment in a Barren Village | −1.020 |
| 7 | New Public Management | 6.65 | Mandarin Training Tutorials | −0.811 |
| 8 | Introduction to Social Security | 6.14 | Leviathan | −0.780 |
| 9 | Modern Administrative | 6.09 | History of Management Thoughts | −0.641 |
| 10 | SPSS Statistical Analysis | 5.98 | The Legend of Lu Xiaofeng | −0.530 |
| 11 | Introduction to Access | 5.91 | Bible Story | −0.250 |
| 12 | Governmental Public Relations | 5.90 | Management Consulting | −0.170 |

**Table 5**. Top 24 Important Books Correlated with Academic Performance Based on $f(\boldsymbol{q}_j)$ of the School of Foreign Language

| Order | Top 12 Positively Correlated Books | | Top 12 Negatively Correlated Books | |
|---|---|---|---|---|
| | Name of Important Books | $f(\boldsymbol{q}_j)$ | Name of Important Books | $f(\boldsymbol{q}_j)$ |
| 1 | Appreciation of English & American Literature | 8.73 | The Protestant Ethic and Spirit of Capitalism | −2.99 |
| 2 | English Interpretation Practice | 8.38 | Watching Anime to Learn Japanese | −2.05 |
| 3 | Emma | 8.22 | The Pursuit of Happiness | −1.99 |
| 4 | What Happens to America | 7.83 | Butterfly Dream | −1.85 |
| 5 | American Literature | 7.80 | The Continuation of Western Culture History | −1.72 |
| 6 | Translation Appreciation and Criticism | 7.51 | Weekly Schedule for Listening (TEM-4) | −1.51 |
| 7 | Japanese Love Story | 7.46 | Latest Guide of Self-Help Abroad | −1.15 |
| 8 | TEM-8 Translation | 7.09 | The Quintessence of Japanese Drama | −1.08 |
| 9 | English Literature Selected Reading | 7.00 | A Tale of Two Cities | −1.08 |
| 10 | Enjoy Chinese Quintessence via English | 6.88 | Dracula's Guest | −0.87 |
| 11 | Inter-Cultural Communication | 6.81 | New Study of British history | −0.76 |
| 12 | Pride and Prejudice | 6.61 | Meteor of the Trip | −0.69 |

as "Artificial Intelligence", "Data Mining" and "Programming Contest Examples", and important major-related books such as "Engineering a Compiler" and "Operation System"; negatively correlated books include engineering-type books, such as "Practical Java Training", and job-preparation books such as "PHP Job Collection", "Mastering Spring", "Introduction to Node.js" and "Game Programming", as well as "Hacker Techniques". For PPM undergraduate students, most of positively correlated books are major-related, such as "Global Civil Society" and "Social Security"; negatively correlated books include extracurricular philosophy books, such as "Interpretation of Hegel's Philosophy" and "Introduction to Psychology", and books implicitly reflecting PPM, such as "Leviathan" and "Bible Story". Since these books may be not useful for their terminal examination, their borrowers may not improve their academic performance. For FL undergraduate students, the positively correlated books share similar characteristic, major-related, such as "English Interpretation Practice", "Translation Appreciation and Criticism", and "Inter-Cultural Communication", while most of negatively correlated books are drama/anime/movie-related books, such as "Watching Anime to Learn Japanese", "The Pursuit of Happiness", and "Dracula's Guest". In other words, lower performance students learn foreign languages by reading drama/anime/movie books, and thus they may not directly improve their academic performance.

### 5.2.4 Library Book Recommendation

We compare the proposed algorithm with three baselines. The first is WMF, without taking book categories and student performance into account; the second one is BPRMF[52], a widely-used recommendation algorithm on implicit feedback datasets; the final one is MostPopular, which recommends books based on the popularity. The comparison results are shown in Fig.6. The observation that WMF outperforms BPRMF indicates the superiority of WMF in library book recommendation based on the book-loan history. By comparing SCWMF-MTL with WMF, we observe the benefit of incorporating book categories and academic performance into the latent factor model. And the superiority of the latent factor models to MostPopular implies what students borrow does not simply depend on the popularity of books. However, the overall recommendation performance is comparatively low. This potentially lies in the extreme sparsity of the student-book loan matrix and a lack of books' external rich information.

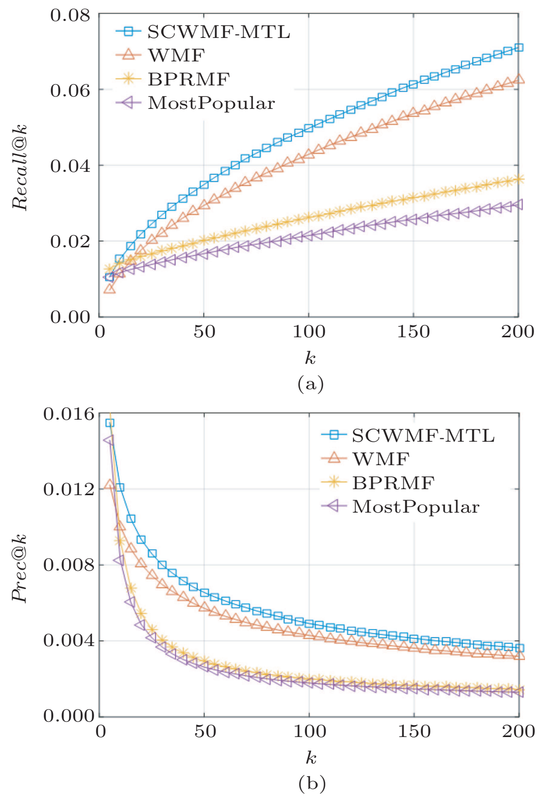



Fig.6. Comparison of recommendation performance. (a) Recall and (b) precision, with baselines.

## 6 Conclusions

In this paper, we studied academic performance prediction based on students' book-loan history, and proposed a supervised dimension reduction algorithm with multi-task learning for collaborative academic performance prediction and library book recommendation. Therefore, these two tasks not only are performed simultaneously but also benefit from each other, as evaluated in the experimental part. According to the analysis to this model, we gave a precise definition of the prediction function and proposed a metric for quantifying the importance of books. Since this prediction function depends on both books borrowed by students themselves and similar books to these borrowed ones, academic performance prediction can be improved based on dimension reduction techniques. We evaluated the proposed model on a dataset of 16 704 students from 14 schools spanning three consecutive grades, and demonstrated the strong effectiveness of the proposed model at academic performance prediction and library book recommendation. In the future, we will consider other external sources of students' behavioral data for further improvement in academic performance prediction and library book recommendation.

## References


[1] Corbett A T, Anderson J R. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 1994, 4(4): 253-278.

[2] Reye J. Student modelling based on belief networks. *International Journal of Artificial Intelligence in Education*, 2004, 14(1): 63-96.

[3] Piech C, Bassen J, Huang J, Ganguli S, Sahami M, Guibas L, Sohl-Dickstein J. Deep knowledge tracing. In *Proc. Annual Conference on Neural Information Processing Systems*, December 2015, pp.505-513.

[4] de Torre J. The generalized DINA model framework. *Psychometrika*, 2011, 76(2): 179-199.

[5] Rasch G. On general laws and the meaning of measurement in psychology. In *Proc. the 4th Berkeley Symposium on Mathematical Statistics and Probability*, July 1960, Volume 4, pp.321-333.

[6] Embretson S E, Reise S P. Item Response Theory. Psychology Press, 2013.

[7] Thai-Nghe N, Horváth T, Schmidt-Thieme L. Factorization models for forecasting student performance. In *Proc. the 4th International Conference on Educational Data Mining*, July 2011, pp.11-20.

[8] Desmarais M C. Mapping question items to skills with non-negative matrix factorization. *ACM SIGKDD Explorations Newsletter*, 2012, 13(2):30-36.

[9] Sun Y, Ye S W, Inoue S Y, Sun Y. Alternating recursive method for Q-matrix learning. In *Proc. the 7th International Conference on Educational Data Mining*, July 2014, pp.14-20.

[10] Töscher A, Jahrer M. Collaborative filtering applied to educational data mining. In *Proc. the KDD Cup 2010 Workshop*, July 2010. http://pslcdatashop.org/KDDCup/workshop/papers/KDDCup2010_Toescher_Jahrer.pdf, June 2018.

[11] Barnes T. The Q-matrix method: Mining student response data for knowledge. In *Proc. AAAI Educational Data Mining Workshop*, July 2005.

[12] Taylor C, Veeramachaneni K, O'Reilly U M. Likely to stop? Predicting stopout in massive open online courses. arXiv:1408.3382, 2014. https://arxiv.org/abs/1408.3382, May 2018.

[13] Halawa S, Greene D, Mitchell J. Dropout prediction in MOOCs using learner activity features. *Experiences and Best Practices in and Around MOOCs*, 2014, 37: 1-10.

[14] Qiu J Z, Tang J, Liu T X, Gong J, Zhang C H, Zhang Q, Xue Y F. Modeling and predicting learning behavior in MOOCs. In *Proc. the 9th ACM International Conference on Web Search and Data Mining*, February 2016, pp.93-102.

[15] Anderson A, Huttenlocher D, Kleinberg J, Leskovec J. Engaging with massive online courses. In *Proc. the 23rd International Conference on World Wide Web*, April 2014, pp.687-698.

[16] Ramesh A, Goldwasser D, Huang B, Daume III H, Getoo L. Learning latent engagement patterns of students in online courses. In *Proc. the 28th AAAI Conference on Artificial Intelligence*, July 2014, pp.1272-1278.

[17] Tamhane A, Ikbal S, Sengupta B, Duggirala M, Appleton J. Predicting student risks through longitudinal analysis. In *Proc. the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2014, pp.1544-1552.

[18] Lakkaraju H, Aguiar E, Shan C, Miller D, Bhanpuri N, Ghani R, Addison K L. A machine learning framework to identify students at risk of adverse academic outcomes. In *Proc. the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2015, pp.1909-1918.

[19] Kotsiantis S, Patriarcheas K, Xenos M. A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems*, 2010, 23(6): 529-535.

[20] Mann P H. Students and Books. Routledge and Kegan Paul, 1974.

[21] de Jager K. Impacts and outcomes: Searching for the most elusive indicators of academic library performance. In *Proc. the 4th Northumbria International Conference on Performance Measurement in Libraries and Information Services*, August 2001, pp.291-297.

[22] Mezick E M. Return on investment: Libraries and student retention. *The Journal of Academic Librarianship*, 2007, 33(5): 561-566.

[23] Goodall D, Pattern D. Academic library non/low use and undergraduate student achievement: A preliminary report of research in progress. *Library Management*, 2011, 32(3): 159-170.

[24] Lian D F, Ye Y Y, Zhu W Y, Liu Q, Xie X, Xiong H. Mutual reinforcement of academic performance prediction and library book recommendation. In *Proc. the 16th International Conference on Data Mining*, December 2016, pp.1023-1028.

[25] Martinez D. Predicting student outcomes using discriminant function analysis. In *Proc. the 39th Annual Meeting of the Research and Planning Group*, May 2001. https://files.eric.ed.gov/fulltext/ED462116.pdf, May 2018.

[26] Thai-Nghe N, Drumond L, Horváth T, Schmidt-Thieme L. Multi-relational factorization models for predicting student performance. In *Proc. KDD Workshop on Knowledge Discovery in Educational Data*, August 2011.

[27] Wu R Z, Liu Q, Liu Y P, Chen E H, Su Y, Chen Z G, Hu G P. Cognitive modelling for predicting examinee performance. In *Proc. the 24th International Joint Conference on Artificial Intelligence*, July 2015, pp.1017-1024.

[28] González-Brenes J P, Mostow J. Dynamic cognitive tracing: Towards unified discovery of student and cognitive models. In *Proc. the 5th International Conference on Educational Data Mining*, Jun 2012, pp.49-56.

[29] He J Z, Bailey J, Rubinstein B I P, Zhang R. Identifying at-risk students in massive open online courses. In *Proc. the 29th AAAI Conference on Artificial Intelligence*, January 2015, pp.1749-1755.

[30] Balakrishnan G. Predicting student retention in massive open online courses using hidden Markov models. Technical Report, University of California, 2013. https://www2.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-109.pdf, May 2018.

[31] Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 2013, 110(15): 5802-5805.

[32] Wu Y Y, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 2015, 112(4): 1036-1040.

[33] Zhong Y, Yuan N J, Zhong W, Zhang F Z, Xie X. You are where you go: Inferring demographic attributes from location check-ins. In *Proc. the 8th ACM International Conference on Web Search and Data Mining*, February 2015, pp.295-304.

[34] Zhu Y, Wang X, Zhong E H, Liu N N, Li H, Yang Q. Discovering spammers in social networks. In *Proc. the 26th AAAI Conference on Artificial Intelligence*, July 2012, pp.171-177.

[35] Culotta A, Ravi N K, Cutler J. Predicting the demographics of Twitter users from website traffic data. In *Proc. the 29th AAAI Conference on Artificial Intelligence and the 27th Innovative Applications of Artificial Intelligence Conference*, January 2015, pp.72-78.

[36] Hu J, Zeng H J, Li H, Niu C, Chen Z. Demographic prediction based on user's browsing behavior. In *Proc. the 16th International Conference on World Wide Web*, May 2007, pp.151-160.

[37] Mooney R J, Roy L. Content-based book recommending using learning for text categorization. In *Proc. the 5th ACM Conference on Digital Libraries*, June 2000, pp.195-204.

[38] Huang Z, Chung W Y, Ong T H, Chen H. A graph-based recommender system for digital library. In *Proc. the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries*, July 2002, pp.65-73.

[39] Noia T D, Cantador I, Ostuni V C. Linked open data-enabled recommender systems: ESWC 2014 challenge on book recommendation. In *Semantic Web Evaluation Challenge*, Presutti V, Stankovic M, Cambria E, Cantador I, Iorio A D, Noia T D, Lange C, Recupero D R, Tordai A (eds.), Springer, 2014, pp.129-143.

[40] Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles. In *Proc. the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2011, pp.448-456.

[41] Wang H, Wang N Y, Yeung D Y. Collaborative deep learning for recommender systems. In *Proc. the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2015, pp.1235-1244.

[42] Lian D F, Ge Y, Zhang F Z, Yuan N J, Xie X, Zhou T, Rui Y. Content-aware collaborative filtering for location recommendation based on human mobility data. In *Proc. IEEE International Conference on Data Mining*, November 2015, pp.261-270.

[43] Zhang Y, Yin H Z, Huang Z, Du X Z, Yang G W, Lian D F. Discrete deep learning for fast content aware recommendation. In *Proc. the 11th ACM International Conference on Web Search and Data Mining*, February 2018, pp.717-726.

[44] Xie M, Yin H Z, Wang H, Xu F J, Chen W T, Wang S. Learning graph-based POI embedding for location-based recommendation. In *Proc. the 25th ACM International Conference on Information and Knowledge Management*, October 2016, pp.15-24.

[45] Lian D F, Zheng K, Ge Y, Cao L B, Chen E H, Xie X. GeoMF++: Scalable location recommendation via joint geographical modeling and matrix factorization. *ACM Transactions on Information Systems*, 2018, 36(3): Article No. 33.

[46] Lian D F, Ge Y, Zhang F Z, Yuan N J, Xie X, Zhou T, Rui Y. Scalable content-aware collaborative filtering for location recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(6): 1122-1135.

[47] Yin H Z, Chen H X, Sun X S, Wang H, Wang Y, Nguyen Q V H. SPTF: A scalable probabilistic tensor factorization model for semantic-aware behavior prediction. In *Proc. IEEE International Conference on Data Mining*, November 2017, pp.585-594.

[48] Zheng B L, Su H, Hua W, Zheng K, Zhou X F, Li G H. Efficient clue-based route search on road networks. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(9): 1846-1859.

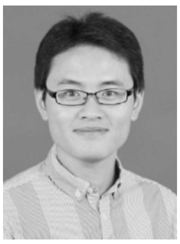
[49] Lian D F, Zhang Z Y, Ge Y, Zhang F Z, Yuan N J, Xie X. Regularized content-aware tensor factorization meets temporal-aware location recommendation. In *Proc. the 16th International Conference on Data Mining*, December 2016, pp.1029-1034.

[50] Hu Y, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. In *Proc. the 8th IEEE International Conference on Data Mining*, December 2008, pp.263-272.

[51] Besag J. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B* (*Methodological*), 1986, B-48(5/6): 259-302.

[52] Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L. BPR: Bayesian personalized ranking from implicit feedback. In *Proc. the 25th Conference on Uncertainty in Artificial Intelligence*, June 2009, pp.452-461.

**De-Fu Lian** is an associate professor in the School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu. He received his B.E. and Ph.D. degrees in computer science from University of Science and Technology of China (USTC), Hefei, in 2009 and 2014, respectively. His research interest includes spatial data mining, recommender system, and learning to hash.



**Qi Liu** is an associate professor at University of Science and Technology of China (USTC), Hefei. He received his Ph.D. degree in computer science from USTC, Hefei, in 2013. His general area of research is data mining and knowledge discovery. He was the recipient of the ICDM 2011 Best Research Paper Award and the Best of SDM 2015 Award.