2016 IEEE 16th International Conference on Data Mining

Mutual Reinforcement of Academic Performance Prediction and Library Book Recommendation

Defu Lian[†], Yuyang Ye[‡], Wenya Zhu[†¶], Qi Liu[‡], Xing Xie[§], Hui Xiong[♯]
[†]Big Data Research Center, University of Electronic Science and Technology of China, dove@uestc.edu.cn
[‡]University of Science and Technology of China, yeyuyang@mail.ustc.edu.cn, qiliuql@ustc.edu.cn
[¶]Hong Kong University of Science and Technology, zhuwenya1991@gmail.com
[§]Microsoft Research, xingx@microsoft.com, [♯]Rutgers University, hxiong@rutgers.edu

***Abstract*—The prediction of academic performance is one of the most important tasks in educational data mining, and has been widely studied in MOOCs and intelligent tutoring systems. Academic performance could be affected with factors like personality, skills, social environment, the use of library books and so on. However, it is still less investigated that how could the use of library books affect 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 promotes library book recommendation by recommending "right" books for students based on their performance levels and book meta information. Finally, we evaluate the proposed model on three years of the 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 is not only predictable from the book-loan history but also improves the recommendation of library books for students.**

I. 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, and notify educators and administrators of in-time intervention, and thus probably prevents delivering the 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], and like cognitive diagnostic models using Deterministic Inputs, Noisy "And" – DINA [2] and item response theory [3]. It is also possible to assign skills to each questions based on (non-negative) matrix factorization [4], [5]. 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 of education. As all students' learning behavior takes place on the Web, based on the recorded data of learning behavior, it is possible for us 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 the factors affecting students' dropout [6], [7], and develop appropriate strategies to retain students in a course. According to previous study [8], [9], [10], behavior like video watching, assignments attempting or quizzes taking, forums posting/viewing/replying and peer friendship, could play an important part 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 [11], [12] based on supervised learning techniques. However, these solutions are not generally applicable in the modern university, 1) because some important data, such as class-attendance records and quiz grades, is rarely digitized, 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, computerized level in the modern university continues to increase, indicating a clear trend for the digitization of students behavior, such as books borrowing and meal having. This makes it possible to predict the future academic performance based on these sources of information.

Library book usage has shown significant contribution to academic success and/or student retention according to past research [13], [14], [15]. This result are further verified by the basic statistics in Fig 1(a), which indicates that students at different performance levels borrow different numbers of books from library. Such statistics analysis is too coarse to answer the questions like "which books could affect students' academic performance". If the relationship of library book usage with academic performance is analyzed in the book level, we would suffer from the sparsity challenge, since each students only borrow 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 predictive analysis of academic

2374-8486/16 $31.00 © 2016 IEEE
DOI 10.1109/ICDM.2016.105

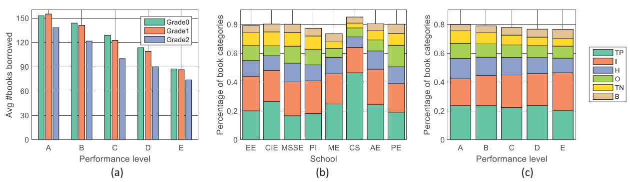

Fig. 1: (a) The averaging book-loan frequency at different performance levels (A(best)-E(worst)); Book category distribution (b) at different schools (e.g. computer science and economic management) and (c) at different performance levels. Category abbreviation refers to Chinese library classification[1]. For example, TP is "Automation, Computer Engineering" and I is "literature".

performance based on the book-loan history and students' academic performance history.

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. Then it feeds them into a regression algorithm with multi-task learning for predicting their academic performance. The reason why to equip 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. Fixed parameters of the regression algorithm, students' book-loan preference is 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 based on our optimization algorithm, 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 alleviate low usage rate of books in modern university library [16]. Based on the latent representation of students, books and book's 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. To the best of our knowledge, this is the first work of jointly modeling book borrowing preference and predicting academic performance.

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 is not only predictable from the book-loan history but also promote the effectiveness of book recommendation.

II. Overview and Preliminary

In this paper, academic performance is predicted based on students' book-loan history. Since each student only borrow a small number of books from library, considering each book as feature index for academic performance prediction would suffer 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, student's 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 [17]. Below, we first make a brief introduction to them.

A. Weighted Matrix Factorization

The proposed algorithm operates on a student-book loan matrix $\mathbf{R} \in \{0,1\}^{M\times N}$, including M students and N books, based on a confidence matrix $\mathbf{W} \in \mathbb{R}^{M\times N}$. Each entry $r_{i,j}$ in the matrix $\mathbf{R}$ indicates whether a student i has borrowed a book j or not, and each $w_{i,j}$ in the matrix $\mathbf{W}$ indicates 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:

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

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

B. 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 [18] should be suggested. This algorithm first represents the profiles of each student by a feature vector $\mathbf{x} \in \mathbb{R}^F$ of F features and represents the content of each book by a feature vector $\mathbf{y} \in \mathbb{R}^L$ 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 $\mathbf{U} \in \mathbb{R}^{F\times K}$ and $\mathbf{V} \in \mathbb{R}^{L\times K}$. Therefore, they can be directly added into latent factors of students and books, i.e., $\mathbf{p}_i = \tilde{\mathbf{p}}_i + \mathbf{U}\mathbf{x}_i$ and $\mathbf{q}_j = \tilde{\mathbf{q}}_j + \mathbf{V}\mathbf{y}_j$. After substituting them into the objecting function (1) and

[1]https://en.wikipedia.org/wiki/Chinese_Library_Classification

regularizing $\mathbf{U}$ and $\mathbf{V}$ by Frobenius norms, content-aware weighted matrix factorization optimizes the objective function as follows:

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

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

III. 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 student's borrowing preference but also absorbs her profile information. This user representation is both able to predict student academic performance by consider them as feature 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 drawing out student preference from borrow 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.

A. 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 them 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 of different schools and their lecturers' teaching skills may be different from each other, they will be eliminated by this preprocessing, making students' CGPA be comparable with each other. Setting the deduced CGPA z_i of each student i as regressand, and considering her latent factor $\mathbf{p}_i$ 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 Fig1(c). For example, students at the school of Humanities and Social Science should benefit from reading novels while students at the school of computer science may not. To address over-fitting problems resulting from data partition by schools, we apply a multi-task learning algorithm for performance regression.

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

$$\mathcal{L} = \sum_i (z_i - \mathbf{e}_i'\mathbf{G}\mathbf{p}_i)^2 + \lambda_D \mathcal{L}_{DR} + \lambda_M \mathrm{tr}(\mathbf{G}'\mathbf{L}\mathbf{G}) + \lambda_R \mathrm{tr}(\mathbf{G}'\mathbf{G}), \tag{3}$$

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

B. Optimization

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

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

$$\mathbf{p}_i = (\frac{\mathbf{g}_s\mathbf{g}_s'}{\lambda_D} + \mathbf{Q}'\mathbf{W}^i\mathbf{Q} + \alpha\mathbf{I}_K)^{-1}(\mathbf{Q}'\mathbf{W}^i\mathbf{r}_i + \alpha\mathbf{U}'\mathbf{x}_i + \frac{z_i\mathbf{g}_s}{\lambda_D}) \tag{4}$$

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

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

$$\mathbf{g}_s = (\mathbf{P}'\mathbf{E}^s\mathbf{P} + \lambda\mathbf{I}_K)^{-1}(\mathbf{P}'\mathbf{E}^s\mathbf{z} + \frac{\lambda_M}{S}(\mathbf{G}'\mathbf{1}_S - \mathbf{g}_s)), \tag{5}$$

where $\lambda = (\lambda_M\frac{S-1}{S} + \lambda_R)$ and $\mathbf{E}^s = \mathrm{diag}(e_{1,s}, \cdots, e_{M,s})$. Due to the existence of Laplacian regularization $\mathrm{tr}(\mathbf{G}'\mathbf{L}\mathbf{G})$, parameters g_s of the prediction model for the school s are affected by other schools, and could play important part in alleviating the over-fitting problem resulting from data partition by schools. Following the 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 $\mathbf{q}_j$ to zero, we can get the updating formula for book latent factor:

$$\mathbf{q}_j = (\mathbf{P}'\mathbf{W}^j\mathbf{P} + \alpha\mathbf{I}_K)^{-1}(\mathbf{P}'\mathbf{W}^j\mathbf{r}_j + \alpha\mathbf{V}'\mathbf{y}_j), \quad (6)$$

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

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

$$\mathbf{U} = (\mathbf{X}'\mathbf{X} + \frac{\beta}{\alpha}\mathbf{I}_F)^{-1}\mathbf{X}'\mathbf{P}, \quad (7)$$

$$\mathbf{V} = (\mathbf{Y}'\mathbf{Y} + \frac{\beta}{\alpha}\mathbf{I}_L)^{-1}\mathbf{Y}'\mathbf{Q} \quad (8)$$

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 $\mathcal{O}(\|\mathbf{X}\|_0 + \|\mathbf{Y}\|_0)K\#iter$ according to [18], 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 alternative least square, that is, taking turns updating 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 $\mathbf{U}$ and $\mathbf{V}$, the complexity of updating $\{\mathbf{p}_i, \mathbf{q}_j, \mathbf{U}, \mathbf{V}\}$ in one round is $\mathcal{O}\big((\|\mathbf{X}\|_0 + \|\mathbf{Y}\|_0)K\#iter + \|\mathbf{R}\|_0 K^2\big)$, since the updating cost for $\mathbf{g}_s$ 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 total number of user-features and item-features.

C. 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, so 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

$$\tilde{z}_i = \mathbf{e}_i'\mathbf{G}(\mathbf{Q}'\mathbf{W}^i\mathbf{Q} + \alpha\mathbf{I}_K)^{-1}(\mathbf{Q}'\mathbf{W}^i\mathbf{r}_i + \alpha\mathbf{U}'\mathbf{x}_i), \quad (9)$$

where α and $\mathbf{W}^i$ follow the same setting as the training phase. Delving into this Eq (9), 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 $\mathbf{Q}'\mathbf{W}^i\mathbf{Q} + \alpha\mathbf{I}_K$, we can rewrite the predictive function as follows:

$$\tilde{z}_i = \mathbf{e}_i'\mathbf{G}\mathbf{Q}'\mathbf{W}^i\mathbf{r}_i/\alpha + \mathbf{e}_i'\mathbf{G}\mathbf{U}'\mathbf{x}_i - \mathbf{e}_i'\mathbf{G}\mathbf{Q}'\big((\alpha\mathbf{W}^{-i} + \mathbf{Q}\mathbf{Q}')^{-1}\mathbf{Q}(\mathbf{Q}'\mathbf{W}^i\mathbf{r}_i/\alpha + \mathbf{U}'\mathbf{x}_i)\big), \quad (10)$$

where $\mathbf{W}^{-i} \triangleq (\mathbf{W}^i)^{-1}$ 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 $\mathbf{e}_i'\mathbf{G}\mathbf{Q}'$. The second type depends on the features of students, weighted by $\mathbf{e}_i'\mathbf{G}\mathbf{U}'$. The third type is the similar books to what they borrow, also weighted by $\mathbf{e}_i'\mathbf{G}\mathbf{Q}'$. The similarity between books is expressed by $(\alpha\mathbf{W}^{-i} + \mathbf{Q}\mathbf{Q}')^{-1}\mathbf{Q}\mathbf{Q}'$. And the final type is the books preferred by a student population having the same features as student i, also weighted by $\mathbf{e}_i'\mathbf{G}\mathbf{Q}'$. 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.

IV. Experiment

The evaluation is conducted on a dataset 16,704 undergraduate students of 19 schools spanning three consecutive grades (denoted as G0, G1 and G2). For each student, this dataset includes 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 classification. In order to learn students' stable borrowing preference, we filter out books which have only been borrowed by less than two students, filter out students who have borrowed less than 5 books and who are at new established schools. The preprocessed dataset includes $13,047$ students from 14 schools. Table I gives the statistics of this dataset.

TABLE I: Statistics of the dataset.

	G0	G1	G2
#students	4,335	4,434	4,278
#books	71,122	72,591	65,183
#records of book-loan	242,376	239,869	194,512
avg. #books borrowed	55.9	54.1	45.5

Based on this dataset, we will evaluate not only academic performance prediction but also 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 Gi), and test it on the dataset of the subsequent grade (Gj), subject to i<j. Denoted as Gi→Gj, 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 first setting of former evaluation corresponds to the cold-start cases, 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.

A. 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 considering the pairwise comparison and measure the ranking consistence within the school s measured by accuracy (abbr. Acc(s)) ,

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

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,

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

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 her borrowing books.

B. Experimental Results

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 the book categories are considered as feature vectors 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 to the proposed models. The comparison results are shown in Table II.

TABLE II: Comparison with baselines

Accuracy	G0→G1	G0→G2	G1→G2	5-CV
Least_MTL	0.5663	0.5696	0.5715	0.5781
WMF_MTL	0.5142	0.5161	0.5249	0.5200
SWMF	0.6138	0.6234	0.6230	0.6335
SCWMF_MTL	0.6279	0.6331	0.6352	0.6438

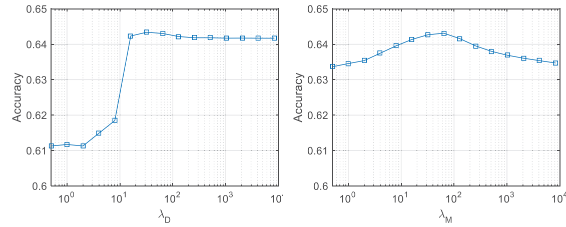

Fig. 2: Sensitivity analysis of λ_M and λ_D in the 5-CV.

From this table, we first observe that Least_MTL is not as 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 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.

To understand the benefit of collaborative learning and multi-task learning, we perform sensitivity analysis of two important parameters, λ_D and λ_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 2, with the increase of λ_D, the performance of the proposed model first improves, and then deteriorate slightly before being stable since this collaborative process is dominated by matrix factorization. The varying trend of Accuracy with the increase of λ_M explicitly shows its optimal value and thus illustrates the effect of multi-task learning once again.

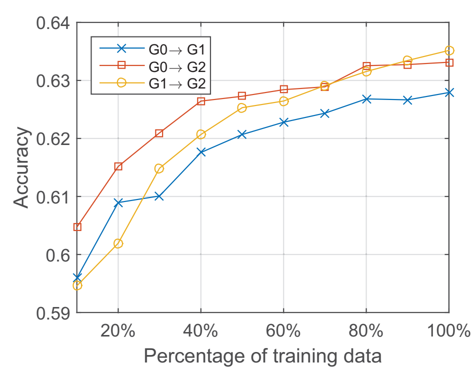

Fig. 3: Impact of data availability.

2) Impact of Amount of Available Data: 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 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 3 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 but with diminishing effect from each additional portion of loan records.

3) Library Book Recommendation: We compare the proposed algorithm with three baselines. The first is *WMF*, without taking book categories and student performance into account; and the second one is *BPRMF* [19], a widely-used recommendation algorithm on implicit feedback datasets; and the final one is *MostPopular*, which recommends books based on the popularity. The comparison results are shown in Fig 4. 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 doesn't simply depend on the popularity of books. However, the overall recommendation performance is comparatively low. This potentially lies in extreme sparsity of the student-book loan matrix and a lack of books' external rich information.

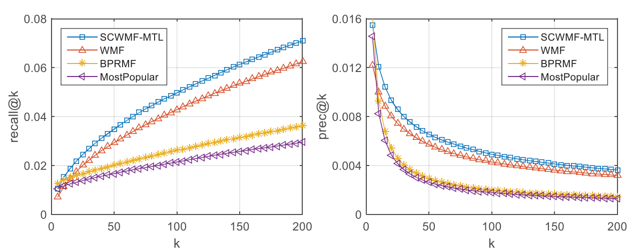

Fig. 4: Recommendation performance comparison.

V. Conclusions

In this paper, we study academic performance prediction based on students' book-loan history, and propose a supervised dimension reduction algorithm with multi-task learning for collaborative academic performance prediction and library book recommendation. Therefore, these two tasks are not only performed simultaneously but also benefit from each other, as evaluated in the experimental part. According to the analysis to this model, we give a precise definition of the prediction function. This prediction function depends on both books borrowed by students themselves and similar books to these borrowed ones, so academic performance prediction can be improved based on dimension reduction techniques. We evaluate the proposed model on a dataset of 16,704 students from tens of schools spanning three consecutive grades, and demonstrate 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' behavior data for further improvement in academic performance prediction and library book recommendation.

Acknowledgement. This work is supported by grants from the Natural Science Foundation of China (61502077, 61672483, 61631005) and the Fundamental Research Funds for the Central Universities (ZYGX2014Z012).

References

[1] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User modeling and user-adapted interaction*, vol. 4, no. 4, pp. 253–278, 1994.

[2] J. De La Torre, "The generalized DINA model framework," *Psychometrika*, vol. 76, no. 2, pp. 179–199, 2011.

[3] G. Rasch, "On general laws and the meaning of measurement in psychology," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 4. University of California Press Berkeley, CA, 1961, pp. 321–333.

[4] N. Thai-Nghe, T. Horvth, and L. Schmidt-Thieme, "Factorization models for forecasting student performance," in *Educational Data Mining 2011*, 2010.

[5] M. C. Desmarais, "Mapping question items to skills with non-negative matrix factorization," *ACM SIGKDD Explorations Newsletter*, vol. 13, no. 2, pp. 30–36, 2012.

[6] S. Halawa, D. Greene, and J. Mitchell, "Dropout prediction in MOOCs using learner activity features," *Experiences and best practices in and around MOOCs*, p. 7, 2014.

[7] J. He, J. Bailey, B. I. Rubinstein, and R. Zhang, "Identifying At-Risk Students in Massive Open Online Courses," in *Proceedings of AAAI'15*, 2015.

[8] J. Qiu, J. Tang, T. X. Liu, J. Gong, C. Zhang, Q. Zhang, and Y. Xue, "Modeling and Predicting Learning Behavior in MOOCs," in *Proceedings of WSDM'16*.

[9] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Engaging with massive online courses," in *Proceedings of WWW'14*. ACM, 2014, pp. 687–698.

[10] A. Ramesh, D. Goldwasser, B. Huang, H. Daume III, and L. Getoor, "Learning latent engagement patterns of students in online courses," in *Proceedings of AAAI'14*, 2014.

[11] A. Tamhane, S. Ikbal, B. Sengupta, M. Duggirala, and J. Appleton, "Predicting student risks through longitudinal analysis," in *Proceedings of KDD'14*. ACM, 2014, pp. 1544–1552.

[12] H. Lakkaraju, E. Aguiar, C. Shan, D. Miller, N. Bhanpuri, R. Ghani, and K. L. Addison, "A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes," in *Proceedings of KDD'15*, 2015.

[13] P. H. Mann, *Students and books*. Routledge and Kegan Paul London, 1974.

[14] K. De Jager, "Impacts and outcomes: searching for the most elusive indicators of academic library performance," in *Proceedings of the 4th Northumbria International Conference on Performance Measurement in Libraries and Information Services*. Association of Research Libraries Washington, DC, 2002, pp. 291–297.

[15] E. M. Mezick, "Return on investment: Libraries and student retention," *The Journal of Academic Librarianship*, vol. 33, no. 5, pp. 561–566, 2007.

[16] D. Goodall and D. Pattern, "Academic library non/low use and undergraduate student achievement: a preliminary report of research in progress," *Library Management*, vol. 32, no. 3, pp. 159–170, 2011.

[17] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Proceedings of ICDM'08*. IEEE, 2008, pp. 263–272.

[18] D. Lian, Y. Ge, F. Zhang, N. J. Yuan, X. Xie, T. Zhou, and Y. Rui, "Content-aware collaborative filtering for location recommendation based on human mobility data," in *Proceedings of ICDM'15*. IEEE, 2015.

[19] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proceedings of UAI'09*. AUAI Press, 2009, pp. 452–461.