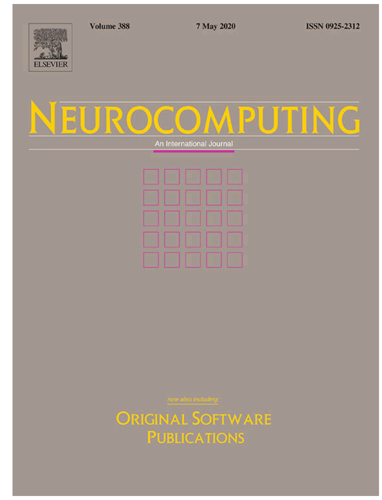
Journal Pre-proofs



Heterogeneous Teaching Evaluation Network Based Offline Course Recommendation with Graph Learning and Tensor Factorization

Yifan Zhu, Hao Lu, Ping Qiu, Kaize Shi, James Chambua, Zhendong Niu

PII: S0925-2312(20)31172-3
DOI: https://doi.org/10.1016/j.neucom.2020.07.064
Reference: NEUCOM 22608

To appear in: *Neurocomputing*

Received Date: 18 February 2020
Accepted Date: 16 July 2020

Please cite this article as: Y. Zhu, H. Lu, P. Qiu, K. Shi, J. Chambua, Z. Niu, Heterogeneous Teaching Evaluation Network Based Offline Course Recommendation with Graph Learning and Tensor Factorization, *Neurocomputing* (2020), doi: https://doi.org/10.1016/j.neucom.2020.07.064

Heterogeneous Teaching Evaluation Network Based Offline Course Recommendation with Graph Learning and Tensor Factorization

Yifan Zhu[a], Hao Lu[a], Ping Qiu[a], Kaize Shi[a], James Chambua[b], Zhendong Niu[a,c,d,*]

[a]*School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China*

[b]*Computer Science and Engineering Department, University of Dar-es-Salaam, Dar es Salaam 35091, Tanzania*

[c]*Library, Beijing Institute of Technology, Beijing 100081, China*

[d]*School of Computing and Information, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Abstract

Course recommendation systems are applied to help students with different needs select courses in a large range of course resources. However, a student's needs are not always determined by their personal interests, they are also influenced by teachers, peers etc. Unlike online courses, user behavior and user satisfaction of offline courses often have serious sparse and cold start issues, which cause overfitting problems in previous neural network and matrix factorization (MF) models. Additionally, the interpersonal relations, evaluation text and existing "user-item" formatted rating matrix constitute a multi-source and multi-modal data structure, so a systematic data fusion method is needed to establish recommendations based on these heterogeneous characteristics. Therefore, a hybrid recommendation model by fusing network structured feature with graph neural networks and user interactive activities with tensor factorization was proposed in this paper. First, a graph structured teaching evaluation net- R2.2

*Corresponding author

Email addresses: zhuyifan@bit.edu.cn (Yifan Zhu), luhao@bit.edu.cn (Hao Lu), qiupcl@163.com (Ping Qiu), kzshi@bit.edu.cn (Kaize Shi), jchambua@gmail.com (James Chambua), zniu@bit.edu.cn (Zhendong Niu)

work is proposed to describe students, courses, and other entities by using the students' rating, commentary text, grading and interpersonal relations. Then, a random walk based neural network is employed to generate the vectorized representation of students by learning their own relational structure. Finally, by recognizing these personalization features as the third dimension of the rating tensor, a Bayesian Probabilistic Tensor Factorization-based tensor factorization is applied to learn and predict students' ratings for classes they have not taken. Experiments on a real-world evaluation of teaching system including 532 participants with 7,453 rating records show that the proposed method outperforms other existing neural network and matrix factorization models including xSVD++, RTTF and DSE with a smaller predictive error as well as better recommendation accuracy.

Keywords: Offline course recommendation, Tensor factorization, Teaching evaluation network, Rating prediction, Personalized recommendation, e-learning.

---

1. Introduction

As a typical type of learning objects and resources, educational courses are an important element in the dissemination of knowledge in online and offline educational scenarios [1, 2, 3]. Currently, both e-learning systems (e.g. Massive Open Online Courses) and conventional teaching departments are suffering from the problem of information overload, that is, a platform often has a large amount of course resources, while students who have just entered the platform do not have a way to quickly and accurately select the most suitable courses due to the lack of overall assessment of self-preferences, skills and learning objectives [4]. Therefore, recommending courses based on the behavior and attribute characteristics of students to overcome such a problem has been the focus of several academic communities [5, 6, 7].

In recent years, although the current course recommendation research has been remarkably focused on both academic and industrial areas as a novel perspective of smart city studies [8, 9, 10]. However, key implied information which helps to improve the accuracy of the recommendation, such as the relations between courses, teachers and students are not fully utilized. Especially in the offline teaching environment, the collection of students' learning information and the extraction of students' relationship network information become more challenging [11]. For instance, the students' choice of courses will not only depend on their own interests and skills, but will also be influenced by suggestions from their peers and teachers. Therefore, an unified representation paradigm for aforementioned text as well as network structure formatted multi-sourced heterogeneous data is significantly required [12]. R2.2

From a technical point of view in the course recommendation area, data from teaching evaluations provide stable information to relieve the above issue, where students submit their ratings and comments to express their opinion on courses they have taken [13]. These ratings build a basic user-course format rating matrix which can be decomposed to student and course vectors to profile students and courses, thereby enabling the ratings of courses which have not yet been taken to be predicted by performing calculations on these vectors [14, 15, 16]. However, traditional matrix factorization (MF)-based recommendation models only consider information such as the user's selection preferences and cannot integrate the above multi-modal heterogeneous network features into the existing rating matrix [17, 18, 19].

Faced with the aforementioned problems and issues, in this paper, we propose a course recommendation model based on an embedding teaching evaluation network and Bayesian Probabilistic Tensor Factorization (BPTF). In particular, we first construct a teaching evaluation network to describe each student's profile relations including academic grade, supervisor, gender and their lexical style in commenting on different courses. Then, the node2vec algorithm is utilized to embed students into vectors, and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is utilized to categorize these vectors into different degrees of personalization. Third, we extend the original rating matrix into a three-dimensional rating tensor which includes students, courses and a category of personalization, and conduct rating prediction based on the BPTF method. We compare and analyze our proposed method with other existing matrix and tensor factorization-based models using a real-world teaching evaluation dataset by selecting the courses with the highest predicted ratings as the recommended list.

In brief, the proposed method has the contributions in the following aspects: (1) First, a heterogeneous graph was constructed for modeling the teaching evaluation with six types of entities and five kinds of relations. (2) Afterwards, by transforming student information, interpersonal relationships, and evaluation text into a network structure, we realized the unified expression of multiple heterogeneous data in the field of course recommendation problems, thereby R2.1 achieving more accurate and effective course recommendations through tensor factorization technologies. To the best of our knowledge, it is the first time that these multimodal and heterogeneous course information have been integrated into a graph based network for depicting students' personalization. Compared to other course recommendation model, our proposed hybrid method has an advantage potential in both rating prediction and top-N recommendation.

2. Related Works

The innovation of the Internet and online e-learning systems have resulted in a huge increase in the volume of course-related information available to students. Early course recommendations have two salient features: using specific rules and using explicit student feedback to make recommendations [20, 21, 22]. With the development of related technologies in the fields of data science and artificial intelligence, and the advancement of modern learning technologies such as intelligent tutor systems, the data sources that can be used for analysis purposes are wider and the analytical approaches are more diverse. Currently, the existing methods for course recommendation can be briefly divided into the following types: collaborative filtering-based, content-based, graph/ontology-based, sequence mining-based, semi-supervised and hybrid methods.

2.1. Collaborative filtering based methods

Collaborative filtering-based methods are widely used in the area of recommendation. These methods make the assumption that if a user has the same opinion as another user on the same item, these two users tend to have more similar opinions on different items than two users chosen randomly. In the area of course recommendation, several studies have focused on collaborative filtering-based methods. Durao et al. built an e-learning system, namely Atepassar.com, and proposed a course-based recommendation system based on calculating the distance between the user matrix and courses [23]. Elbadrawy and Karypis introduced matrix factorization into course recommendation and proposed a neighborhood-based user collaborative filtering method to generate top-n course recommendation [24]. This MF-based method was enhanced by Thanh-Nhan et al. who implemented a k-nearest neighborhood (kNN) collaborative filtering method to improve the performance of the matrix factorization [25]. Khorasani et al. proposed a Markov-based collaborative filtering model to recommend courses to students each semester, based on the sequence of courses they have taken in the previous semesters without any prior knowledge of the institution, course prerequisites, curriculum or degree requirement [26]. In addition, two-step collaborative filtering based method is proposed by Lee et al. [27]. This study divided Bayesian Personal Ranking Matrix Factorization into a two-stage algorithm to solve the problem where students of different grades are different in the choice of public electives. Recently, the Probability Matrix Factorization (PMF) model was introduced into the course recommendation scenario by Li et al. [28]. Huang et al. proposed a Cross-User-Domain collaborative filtering method using the course score distribution of the most similar senior students to achieve a better prediction performance [29]. Recently, Lin et al. presented a convex optimization-based framework with one L0 regularization and the constraint on the learners' characteristics. It leveraged a machine-learning process to automatically extract formatted and un-formatted data consisting of both global and local features of the courses [30]. R1.6

2.2. Content based methods

Content based methods are based on a description of the recommended item and a profile of the user's preferences. Apaza et al. utilized Latent Dirichlet Allocation (LDA) to calculate the topic of courses, thereby recommending them to students with similar topic preferences [31]. To understand the features of a course better, a deep course content analysis method is performed. Dai et al. also employed LDA to predict how the content of a course is distributed over different categories in the domain, after which they utilized the syllabus as course content and curriculum guidelines as domain knowledge to build the recommender system [32].

2.3. Graph based methods

As a learning resource object, courses in a cyber and physical educational system contain a series of entities and relations, which can constitute an entity-relations network of the graph structure. Also known as a knowledge graph and ontology, graph based approaches can make efficient use of this natural data structure advantage [33]. Antony et al. proposed an short message service based survey program to collect students' opinion on courses, and combined these opinions with their pre-defined course ontologies to support fuzzy-based course recommendation [34]. An ontology-based course recommender application is deployed by Ibrahim et al. with a more detailed definition of course ontology [35]. Their study and application are further refined by the kNN algorithm which is reported in [36].

2.4. Sequence mining based methods

In addition, several studies focus on the use of each student's course sequence for sequence mining to obtain recommendations. Aher and Lobo exploited an Apriori+kmeans method to recommend courses in an e-Learning system based on historical data [37]. Xu et al. exploited a forward-search backward-induction algorithm to optimally select a course sequence and used multi-armed bandits to optimally recommend a course sequence [38]. Zhang et al. implemented the classical Apriori algorithm to mine a frequent itemset in a Hadoop parallel computing environment [39]. Hou et al. also proposed a novel online learning algorithm based on hierarchical bandits with known smoothness to face the problems of the heterogeneity of large-scale user groups, sequence problems in courses and the foreseeable quantitative explosion of courses and users [40]. Based on the Markkov chain framework, Polyzou et al. proposed a random-walk-based approach that captures the sequential relations between the different courses and recommended a short list of courses using a crowdsourcing approach [41].

2.5. Semi-supervised methods

Recommending courses to students does not necessarily require the use of R2.4
fully supervised methods. For example, Aditya et al. gave a unique insight from the linear program which their goal was to recommend courses to students that not only satisfy constraints, but that are also desirable [42]. As another example, Hoiles and van der Schaar gave a novel view from Reinforcement Learning. They proposed a regression estimator for contextual multi-armed bandits with a penalized variance term [5]. In addition, a taxonomy-based solution inspired by the domain of library science was proposed by Yang et al. [43]. Recently, based on the advantages of reinforcement learning, Zhang et al. proposed a Hierarchical Reinforcement Learning method for mass user learning data on their e-learning platform [7].

2.6. Hybrid methods

Several studies propose a recommendation approach by combining at least two of the aforementioned methods. Hybrid methods have received attention since the beginning of research into course recommendation. For example, in earlier periods, scholars used collaborative filtering as a complement to the rule-based method to reduce invalid recommendations due to missing rules [44]. Additionally, the collaborative filtering method can also be improved by the immune algorithm which was reported by Chang et al. [45]. Ng et al. proposed a college-oriented course recommendation system, namely CrsRecs which combines LDA, tag and sentiment methods to extract students' opinions on courses and uses them with the matrix factorization method to make recommendations [46]. Jing et al. [6] also utilized "collaborative filtering + content based" methods to recommend courses to relieve the issue of sparsity, anti-interest and the cold start problem. Lin et al. proposed a sparse linear method based top-n course recommendation system with artificial expert knowledge [47]. Krstova et al. made a combination of SimRank and matrix factorization method to implement course recommendations. Recently, Esteban et al. [48] presented a Genetic Algorithm based recommendation system configuration which combines collaborative filtering and content-based filtering by using multiple criteria related both R1.6 to student and course information. They performed an experimental study by using real-world data from University of Cordoba to exhibit the practicability of their proposed model. Additionally, our previous studies also proposed several hybrid models to recommend learning objects including courses based on concept mapping and the immune algorithm [49], sequence pattern mining and ontology [50], as well as collaborative filtering and self-organization-based knowledge propagation network [51] with fuzzy rules enhanced solution [2].

To summarize, the current research on course recommendation utilizes different features from students' historical behaviour, review content, sequence characteristics, ontologies, etc. Based on the materials used in these studies and their results, we can assert that the behavior of student in selecting courses is related to data with multiple modalities from multiple channels. Therefore, seeking a uniform quantitative representation of these multi-source heterogeneous data and relying on as many data features as possible will help to improve the accuracy of the recommendation. However, to the best of our knowledge, there are few solutions which investigate merging the network structured student profile (graph feature), review style (content feature) and student similarity (collaborative filtering feature).

3. Methodology

In this section, we introduce our proposed method for rating prediction and course recommendation. Our proposed method consists of two major parts: (1) measuring the degree of similarity of a student with other students when evaluating a course according to their teaching evaluation network as well as the level of the student's mastery of knowledge of courses, thereby categorizing them into different classes; (2) proposing a rating prediction method by applying BPTF to the student evaluation of teaching (SET) tenor which consists of ratings, students, courses, and students' personalization measured as in (1). The entire schema of this study is presented in Figure 1.

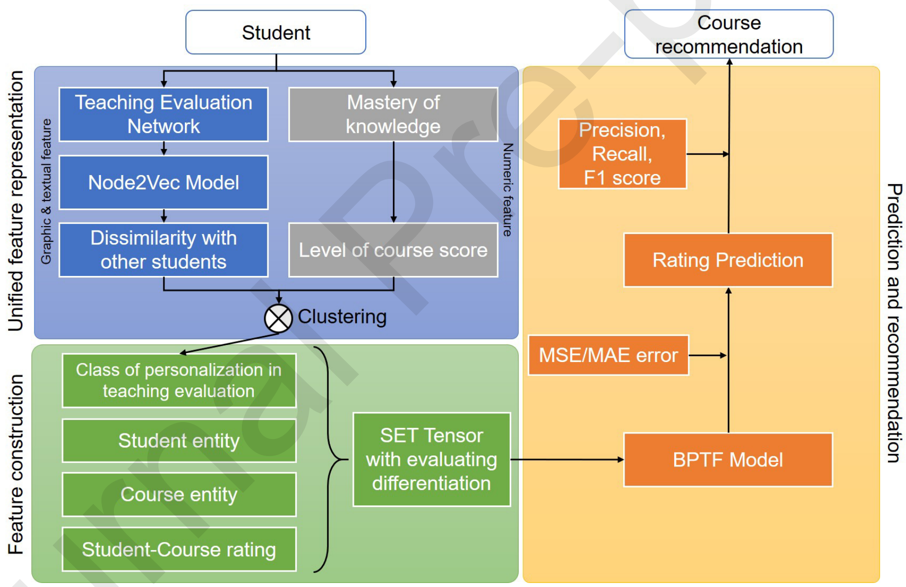


Figure 1: The schema of our proposed method.

3.1. Unified representation of students' personalization in teaching evaluation

3.1.1. Establishing heterogeneous student evaluation of teaching network R1.2

For student $s_i(i = 1, 2, ..., m)$ and course $c_j(j = 1, 2, ...n)$, the corresponding rating record $(re_{i,j})$ can be described as a quadruple:

$$re_{i,j} :=< s_i, c_j, r_{i,j}, mk_{i,j} > \quad (1)$$

where $r_{i,j}$ denotes the review/rating the student gives to the course and $mk_{i,j}$ is the teaching result (e.g. final grades) the student achieves in this course. In this study, to avoid the introduction of potential retaliatory ratings, we adopt a static game-like assumption where students can only view their grades after completing the evaluation of the course. On the other side, the teacher can only view the students' evaluation of the courses they have taught after uploading the students' grades.

Based on the precondition mentioned in the introduction, $r_{i,j}$ is made up of text-based comments (rt) as well as score-based evaluation questionnaires (rs), i.e.

$$r_{i,j} :=< rt_{i,j}, rs_{i,j} > \quad (2)$$

In this paper, we further extract lexical features from textual comments. Based on a previous finding [13], adjectives, verbs, and nouns after the removal of stop words reflect more commentary themes, therefore after the comment sentences are processed through word segmentation and the removal of stop words, only adjectives, verbs, and nouns are saved as nodes in the review network. In this way, the original student-course one-on-one evaluation can be transformed into a student-keyword co-occurrence-course network structure. In other words, the original relation $< s_i, c_j, rt_{i,j} >$ is split into:

$$< s_i, c_j, rt_{i,j} > \Rightarrow \{< s_i, w_k, choose\_word >\}, \{< w_k, c_j, describe >\} \quad (3)$$

where w_k is each theme word in all theme words $W_{i,j}$ given by s_i to c_j, and $W = \bigcup_{i=1,j=1}^{m,n} W_{i,j}$.

From another aspect, there are also structured relations between students outside of this field of teaching evaluation. These relations have the potential to affect a student's course choice, because it has been proven that the course selection preferences are not expected to be independent from others choices, as students may follow trends of similar behaviors [1]. Additionally, the influence R2.5 on the choices is not only made by peers, but also their supervisors because the supervisors tend to advise students to choose courses that are relevant to their research field so that they can lay the foundation for future research work more quickly [52]. For example, different graduate students directed by the same supervisor often choose similar courses considering the research interests of the supervisor. Thus, three additional relations are introduced to construct the teaching evaluation network:

- gender: the gender relation is used between students and entity male/female;
- supervision: the supervision relation is used between students and their supervisors.
- grade: the grade relation is used between students and their corresponding grade.

Combining the two comment-based relations and the direct review relations mentioned above, we present an illustration of these six intrinsic relations and different entities in the teaching evaluation network as shown in Figure 2. In short, different students can be reflected in the teaching evaluation network because of the differences and commonalities in relationships with supervisors, R2.2 grades, genders, comment styles and profile of courses.

3.1.2. Generating embedded representation of student entity in the SET network R1.2

To quantitatively measure the differences between each student, a graph-based presentation method is utilized to vectorize each student entity into numeric vectors.

Inspired by the word embedding model word2vec (specifically the skip-gram model) [53], DeepWalk uses the co-occurrence between nodes in the graph to

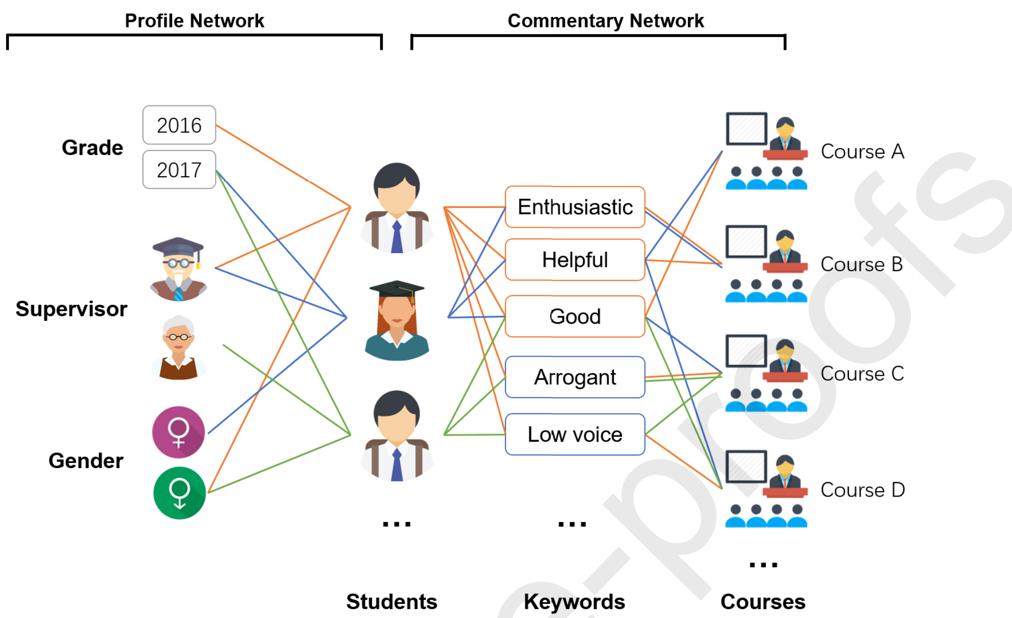


Figure 2: An illustration of the teaching evaluation network. The lines with different colours between the courses and the students represent the courses directly evaluated by each student.

learn the vector representation [54]. A RandomWalk algorithm is utilized by DeepWalk to sample the nodes in the graph, thereby generating access sequences which work like a word sequence in sentences. Then, the skip-gram model can be applied to generate the vector representation of nodes in a similar way to its original application on words. However, when generating a node sequence, the choice of the next node from the currently resident node is determined by a uniform random sampling on all possible edges. This operation neglects the local structural properties and similarity between the adjacent nodes in a graph while these two attributes can be derived by a classical breadth-first search and depth-first search. Based on this, node2vec develops a flexible biased random walk procedure that can explore neighborhoods in a breadth-first search as well as a depth-first search fashion [55].

Let $G = (V, E)$ be the teaching evaluation network, where V denotes the set of all named objects (i.e. entities) including grade, supervisor, gender, student, keyword, and course. E denotes the set of all relations between the aforementioned entities. Given a node u, let c_i denote the i^{th} node in the walk with $c_0 = u$. Nodes c_i are generated by the distribution:

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{1}{Z}\omega_{v,x} & (v,x) \in E \\ 0 & \text{Otherwise} \end{cases} \tag{4}$$

Note that $\omega_{v,x}$ is the unnormalized transition probability between nodes v and x, and Z is the normalizing constant. A 2^{nd} order random walk method with two hyperparameters p and q is selected to guide the walk: Assume a random walk resides at node v which traversed through edge (t, v) from node t, the next edge the walk will traverse is decided on the transition probabilities $\omega_{v,x}$ on edges (v, x) leading from v, which is defined by:

$$\omega_{v,x} = \alpha_{p,q}(t,x) = \begin{cases} \frac{1}{p} & d_{tx} = 0, (v,x) \in E \\ 1 & d_{tx} = 1, (v,x) \in E \\ \frac{1}{q} & d_{tx} = 2, (v,x) \in E \\ 0 & \text{Otherwise} \end{cases} \tag{5}$$

where d_{tx} denotes the shortest path distance between nodes t and x. The hyperparameter p controls the probability of repeatedly accessing the node that has just been accessed, while q controls whether the walk is outward or inward. If $q > 1$, the random walk tends to access nodes close to t (biased to breadth-first search), and if $q < 1$, the walk tends to access nodes away from t (biased to depth-first search). For the sake of simplicity, we use the symbol Ns to represent the aforementioned strategy for walking.

Let d denote the dimension of each entity in the teaching evaluation network and let $f(x)$ denote the mapping from node (i.e. entity) to the corresponding feature representation. For every source node $u \in V$, we define $nNs(u) \subset V$ as a network neighborhood of node u generated through neighborhood sampling strategy Ns. By optimizing the objective function in Eq.6, d-dimensional vector representation of entities especially students can be derived by following:

$$\max_{f} \sum_{u \in V} \left[-\log Z_u + \sum_{n_i \in nNs(u)} f(n_i) \cdot f(u) \right] \tag{6}$$

where Z_u is theoretically defined by $Z_u = \sum_{v \in V} \exp(f(u) \cdot f(v))$, and estimated using negative sampling due to its expensive computational consumption [56].

3.1.3. Measuring students' personalization via clustering R1.2

Combined with the level of the students' mastery of knowledge (i.e. final grades student get from the course) in each class, we transfer the student's review style as well as the effectiveness of learning to a general quantitative similarity by calculating the cosine distance between any two students with a corresponding evaluation score to courses:

$$Sim((s_i, c_k), (s_j, c_l)) = \frac{(f(s_i) \bowtie \beta rs_{s_i,c_k})^T \cdot (f(s_j) \bowtie \beta rs_{s_j,c_l})}{\|f(s_i) \bowtie \beta rs_{s_i,c_k}\| \cdot \|f(s_j) \bowtie \beta rs_{s_j,c_l}\|} \tag{7}$$

where $\bowtie$ denotes the splicing operation of vectors and β is a weighted constraint. Hence, we transfer all the student-course evaluation pairs to evaluation vectors $f(s_i) \bowtie \beta rs_{s_i,c_k}$, and further apply the DBSCAN method [57] to ensure evaluations of a similar style are regarded as the same type (the type is denoted as $t(s_i, c_k)$). In practice, we modify the thresholds to set the types with very few vectors to a default type to reduce the sparsity of the evaluation. To summarize, the process of the proposed unified feature representation requires node2vec which occupied the most computational time complexity. Thus the time complexity of the this part is $O(rul)$, where r is the number of walks per node, u is the number of nodes and l is the walk length. R2.6

3.2. Feature construction with SET Tensor R1.2

After assessing the students' personalization in teaching evaluation, we transfer the original heterogeneous quadruple into a quantitative one:

$$re^{'}_{i,j} := < s_i, c_j, t(s_i, c_j), rs_{i,j} > \tag{8}$$

Combining the above $re^{'}$ in three dimensions: student, course, and personalized type, we get a three-dimensional tensor, namely the SET tensor. Compared with the "user-item" rating matrix for probabilistic matrix factorization [58] and other matrix factorization methods, the SET tensor introduces review preferences and structural differences as a third dimension to extend the feature space, thereby reforming the original hypothesis that user vectors and item vectors are independent and identically distributed [59]. In other words, the introduction of the SET tensor solves a problem not previously considered by the MF-based methods, which is a contribution that can be made by correctly modeling the user's profile and reviewing text. By using the first three elements R2.2
of quadruple $re^{'}_{i,j}$ as a dimension and the last dimension as a value, We construct the SET tensor to prepare the data for evaluation prediction and further recommendation. Hence, the traditional rating matrix can be arranged in a three-dimensional tensor whose three dimensions correspond to student, course, and user personalization type with sizes m, n, and K, respectively ($R^{m \times n \times K}$), which can be expressed as:

$$R^k_{i,j} \approx < S_i, C_j, T_k >= \sum_{d=1}^{D} S_{di} C_{dj} T_{dk} \tag{9}$$

where S_i, C_j and T_k are D-dimensional vectors decomposed from the SET tensor and rsi, j^k is the rating rs_{s_i,c_j} from the student s_i to course c_j under a degree of personalization $t(s_i, c_j)$. In addition, Eq. 9 can be interpreted as that a rating depends not only on the student and course, but also the classroom performance, semantic meaning in the review text, as well as the personal relations with colleagues, supervisor, etc. Additionally, in a probabilistic model, the issue of randomness should be considered. The conditional distribution $R^k_{i,j}$ given by U, V, T is a Gaussian distribution with mean $< S_i, C_j, T_k >$ and variance α^{-1}, i.e.

$$R^k_{i,j} | S, C, T \sim N(< S_i, C_j, T_k >, \alpha^{-1}) \tag{10}$$

For example, under the extreme condition that all users' personalization are very similar, T_k in Eq.10 are all-one vectors and the model is equivalent to the common PMF model. However, this condition may cause overfitting due to the sparseness of the rating tensor (missing entries in R). Therefore, to avoid the overfitting problem caused by rating tensor R, we apply a Bayesian treatment to the PMF model based on our introduced similarity of personalization. This is derived by imposing a zero mean and independent Gaussian priors on the entities' feature vectors, which is proposed by [60]:

$$S_i \sim N(0, \sigma_S^2 \boldsymbol{I}) \quad i = 1, 2, ..., m \tag{11}$$

$$C_j \sim N(0, \sigma_C^2 \boldsymbol{I}) \quad j = 1, 2, ..., n \tag{12}$$

where $\boldsymbol{I}$is D-by-D identity matrix.

By adding structural and semantic similarity (i.e. similarity of personalization) between the students' reviews as an additional factor, we hypothesize that different structural and semantic similarity has a different influence on teaching evaluation scores. On this basis, we assume each similarity of a particular student review on a course is conditioned on that the similarity of the remaining reviews on the same course excluding this one. In other words, the applied conditional prior is as follows:

$$T_k \sim N(T_k^{'}, \sigma_{dT}^2 \boldsymbol{I}) \quad i = 1, 2, ..., K \tag{13}$$

295 where $T_k^{'}$ is the personal similarity of the reviews excluding the k^{th} review.

3.3. BPTF-based rating prediction and recommendation R1.2

Based on the above constraints, the stochastic gradient descent method (SGD) is applied to learn the model parameters automatically with regularization and is combined with the sparse observations to estimate the unknown ratings by maximizing the logarithm of posterior distribution. This is equal to minimizing the following loss function:

$$\begin{aligned} L(S, C, T) = &\sum_i^m \sum_j^n \sum_k^K \boldsymbol{I}_{i,j}^k (R_{i,j}^k - \langle S_i, C_j, T_k \rangle)^2 \\ &+ \sum_{i=1}^m \frac{\lambda_S S_i^2}{2} + \sum_{j=1}^n \frac{\lambda_C C_i^2}{2} + \sum_{k=1}^K \frac{\lambda_{dT} T_k - T_k^{'2}}{2} \end{aligned} \tag{14}$$

where $\lambda_S = (\alpha\sigma_S^2)^{-1}, \lambda_C = (\alpha\sigma_C^2)^{-1}$ and $\lambda_{dT} = (\alpha\sigma_{dT}^2)^{-1}$.

Finding only one point to estimate the parameters of the learning model instead of inferring the full posterior distribution largely contributes to the efficiency of training a tensor-based PMF model [61]. However, the maximum a posteriori scheme which was utilized to compute such point estimates often suffers from overfitting if hyper parameters are not set properly. The problem becomes more probable when the dataset is relatively large but sparse. Given a dataset D and parameters θ, the MAP finds the parameter estimates $\theta*$ which maximizes $P(D|\theta)$, and finally the model predicts unseen samples based on such learnt $\theta*$. Such estimation is sensitive to random variations of data which may lead to overfitting [15]. Therefore, in this paper, a complete Bayesian treatment is applied to our model to avoid overfitting. The treatment establishes priors on parameters, and thus proper averages over different models and eventually guarantees the fine tuning of parameters. In particular, Markov Chain Monte Carlo is used as sampling-based approximation method [60]. To summarize, the process of the proposed unified feature representation requires node2vec which occupied the most computational time complexity. Thus the time complexity of the this part is $O(MNK)$. R2.6

4. Experimental results

4.1. Dataset

We perform our experimental evaluation on a real-world data set from the Graduate School, Beijing Institute of Technology which is collected from a graduate teaching evaluation system named BIT-UASET [13]. The online system requires graduate students to review and evaluate each course they have taken each semester. The reviews of the participating students are anonymously shown to the teachers as well as the Department of Teaching Quality Control. In order to avoid irresponsible and random scoring, the system adopts a bidirectional static game-based scoring rule where students can only check their final grades after completing the teaching evaluation, and the teacher can only view the students' evaluation after submitting the students' final scores. The whole dataset consists of student IDs, course IDs, course names, scored ratings, final grades, and text reviews. It should be noted that these scores are between 0-100, which is different from the common five-level scale scheme. To collect information to construct the profile network illustrated in Figure 2, we obtain the student's grade, supervisor and gender from the School of Computer Science and Technology, Beijing Institute of Technology to complete whole teaching evaluation network we designed. Eventually, by selecting the rating records without missing data in the School of Computer Science and Technology during the academic R1.4 year of 2017-2018 and 2018-2019, a total of 7,453 rating records from 532 students in relation to 201 courses are extracted as the raw dataset. R3.2

4.2. Experiment setup and preprocessing of data

4.2.1. Preprocessing teaching evaluation data

We further filter the rating records in the dataset for the purpose of reducing the sparsity of the rating tensor. In the experiment, we only keep students who reviewed at least 5 courses throughout the academic year and the courses which had at least 10 ratings. Under these constraints, there are 7,453 records from 706 students in relation to 76 courses which are utilized in the experiment. A train-test approach is utilized in the following experiment. Therefore, we R1.4 randomly sample 20% students (i.e. 141 students) from the dataset as the R3.2 test set. For the 141 students in the test set, we used the last five courses evaluated by each person as the evaluation labels for rating prediction and course recommendations. Thus, there are 565 students with 5,842 ratings in the training set and 141 students with 1,611 ratings (in which 705 ratings are used for measuring the recommendation performance) in the test set (Table 1)

In terms of lexical features, the students' textual reviews with stop words removed are segmented by the Jieba word segmentation tool [1]. Then, we extracted all theme and sentiment-related content words according to our previously pro-

[1] https://github.com/fxsjy/jieba

Table 1: Description of the dataset.

Dataset	Number of students	Number of ratings
Training set	565	5,842
Test set	141	1,611 (705 used for evaluation)
Total:	706	7,453

R1.4

posed method [13]. Finally, a total of 5,001 words including 2909 theme-related words and 2,092 sentiment-related words are extracted to construct the teaching evaluation network.

4.2.2. Implementation of the proposed method

For the sake of simplicity, we name the proposed model as TENFT (Teaching Evaluation Network-Tensor Factorization) and use this abbreviation in the rest of this paper. The entire experiment is performed on a computer with an Intel i7-7700K CPU, 16GB of memory and a Nvidia GTX1080Ti GPU. As discussed in the previous section, we use node2vec to embed the teaching evaluation network. In this experiment, *OpenNE*[2] is utilized to establish the embedding process. During the training process, the hyper-parameters of node2vec are set as follows: The number of random walks started at the same node is 10, the length of random walk started at each node is 50, the window size of skip-gram model is 5 and the p and q are all equal to 1.0. As a result, we get the embedding of the entities in the teaching evaluation network with a dimension size of 200. Furthermore, we implement SET tensor factorization by modifying Candecomp-Parafac factorization [62] function so that it is compatible with the missing values contained in the rating tensor based on Tensorly framework[63]. The learning rate is set to 0.001. Additionally, we tested different number of latent features to evaluate model behavior, and determine this parameter according to the optimal performance. R1.3

[2]https://github.com/thunlp/OpenNE

4.3. Evaluation metric and baselines

To valid the feasibility and performance of the proposed model, we compare our model with the following methods:

- SVD++ [64]: SVD++ is a classic rating prediction model which introduces an implicit feedback mechanism on the basis of Singular Value Decomposition. It uses the user's historical browsing data, user historical rating data, item's historical browsing data, and item's historical rating data as new implicit parameters to be passed during matrix factorization. In this experiment, we trained the SVD++ model with 30 epochs with learning rate $\gamma = 0.004$ and regularization parameter $\lambda = 0.13$.

- RTTF [15]: RTTF model considers the linguistic similarity between review texts as an additional factor in rating prediction. It establishes the rating prediction based on tensor factorization which involves review text semantic similarity. Their proposed tensor factorization model supplements the central task of factorization methods of finding similar users, uncovering underlying characteristics of the data and predicting user preferences by introducing text semantic similarity. In this experiment, we set the model R1.3 with the recommended hyper-parameter that Wishart $\tilde{W}_0 = 2$, learning rate $\gamma = 0.001$ and regularization parameter $\lambda = 0.015$.

- xSVD++ [1]: The xSVD++ focuses on user's preferences which are not expected to be independent from others choices on the online educational platform. The proposed xSVD++ is a multi-dimensional Matrix Factorization model combined with collaborative filtering algorithm, which exploits information from external courses' characteristics to predict course trends and to perform rating predictions. In this experiment, we set the model with the recommended hyper-parameter that $\gamma = 0.001$ and regularization parameter $\lambda = 0.1$.

- DSE [16]: DSE is a hybrid approach to learn and represent users' preference knowledge from review texts and utilize the acquired representation to support rating prediction. The DSE utilizes Long Short-Term Memory architecture to learn users' preference knowledge along with item aspects and the PV-DM model to convert the acquired knowledge to embedded representation, and establish the rating prediction by aggregating such knowledge through PMF model. In this experiment, we set the model with the original recommended hyper-parameter that learning rate = 0.1 (halved after every third epoch), batch size = 128, source and target sequence length equal to 300 and 30, respectively.

Additionally, to valid the effectiveness of the proposed SET Tensor, we further compare the performance between the proposed TENTF model and its variants which eliminated certain type of feature:

- TENTF-Profile: This model deletes the Profile Network described in section 3.1, leaving only the Commentary Network for the further network embedding.
- TENTF-Commentary: Contrary to the previous item, the TENTF-Commentary deletes the Commentary Network, while leaving only the Profile Network for the further network embedding. R3.1
- TENTF-Mastery: This model removes the numeric feature of the final grades student get from the course.

Considering that the intent of this paper is to use previous evaluation results to predict the students' reviewing rate on a course they have not yet taken (i.e. a rating prediction task), the commonly used root mean squared error (RMSE, Eq.15) and mean absolute error (MAE, Eq. 16) are computed to measure the error between the predicted rating and the real one for the different models.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{15}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{16}$$

where $\hat{y}$ is the predicted rating and y is the actual rating in the test set.

Additionally, course recommendation can be generated by the predicted ratings. In this paper, we use the top-N recommendation scenario to measure the performance of the different models. In particular, we select the ten largest ratings of each student in the test set as the labels, then we measure the Precision@N (Eq.17), Recall@N (Eq.18) and F1@N (Eq.19) between the top N predicted courses which have the largest predicted ratings (N=1,2,3,4,5).

$$\text{Precision} = \frac{\text{Number of courses in the label}}{\text{Number of recommended courses}} \tag{17}$$

$$\text{Recall} = \frac{\text{Number of courses in the label}}{\text{Number of all courses in the label}} \tag{18}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

4.4. Results

Table 2: RMSE and MAE comparison on the test set between the proposed method and the baseline models.

Model	MAE	RMSE
SVD++	11.765	14.3263
xSVD++	8.4237	11.9673
DSE	7.4385	10.7338
RTTF	7.8012	10.9902
TENTF-Profile	8.3584	11.1342
TENTF-Commentary	8.8525	11.9987
TENTF-Mastery	7.2001	9.9017
TENTF (Proposed)	**6.9133**	**9.5913**

R1.4
R3.1
R3.2

As previously discussed, we split the BIT-UASET data into a training set and a test set. The ratings in the test set are marked as a missing value in the SET tensor. We compare the RMSE and MAE between the predicted rating and the real rating for the different models, and the performance of the rating prediction for educational course recommendation is shown in Table 2 and the influence of a different selection of latent factors is shown in Figure 3. It can be observed that our proposed model achieves the smallest prediction error(MAE: 6.9133, RMSE: 9.9017) compared with other PMF inspired approaches (RTTF and DSE) and SVD based methods (xSVD++ and SVD++). Additionally, by eliminating different feature parts, the performance of our proposed model has declined to varying degrees, which shows that our proposed features provide positive contributions to the realization of the final prediction. It can be observed that the personal profile network and commentary network both play a significant role in the rating prediction of the model with a performance loss at 16.10% and 25.09%, respectively. However, the performance loss caused the student's course grades is relatively minor (3.23%).

R1.4
R3.1
R3.2

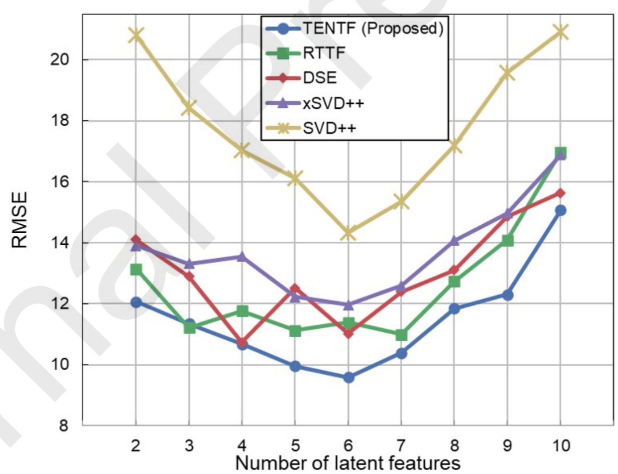


Figure 3: RMSE for different number of factors

As described in the experimental setup, we further explored the impact of different number of latent factors on the predicitng errors. As shown in Figure 3, the model compared in the experiment obtains the smallest prediction error when D is between 5 and 7. In particular, the proposed model derives the smallest prediction error when the number of the latent factors D equals to 6.

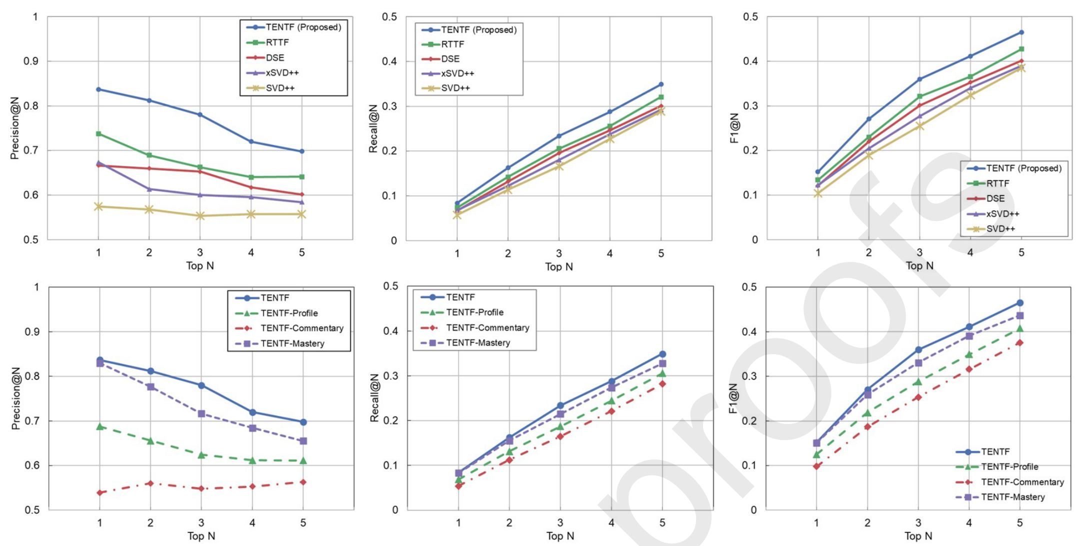


Figure 4: Comparing the performance between our model and the baseline at top-N recommended courses in term of Precision@N, Recall@N and F1@N. R1.4 R3.1 R3.2

Furthermore, by converting the course corresponding to the largest top-N prediction ratings into a recommendation list, we measure the degree of accuracy with the list of courses produced by the actual ratings. As shown in Figure 4, our proposed model has outperformed other baselines as well as its own variants, which derived the optimal performance in Precision@N , Recall@N and F1@N. In particular, the proposed model can keep the precision rate almost above 70% when N changes from 1 to 5 (Precision@5 = 0.6978), which is superior to other baseline models. On the other side, the recall indicator has shown that the proposed model derived the optimal performance (Recall@5 = 0.3489) compared with baseline models. In terms of the variants of the proposed model, the results are similar to the rating prediction task, indicating our proposed features all have contribution to the recommendation model. The result also indicates that the removal of students' profiling or commentary information leads major performance loss to the current model while eliminating the students' mastery of knowledge has a minor contribution to the model.

5. Discussion

Based on the above experimental results, the following points are worth noting and discussion.

First, it should be noted that the objective function is not unique. According to previous studies, a stochastic gradient ascent method can be utilized to replace the SGD which is used in our experiments with the following objective function:

$$\begin{aligned}
Objective =& \log p(S,C,T|R) \propto \log p(R|S,C,T) + \log p(S,C,T) \\
=& \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{K} \boldsymbol{I}_{i,j}^{k} \log p(R_{i,j}^{k}|S_i,C_j,T_k) \\
&+ \sum_{i=1}^{m}\log p(S_i) + \sum_{j=1}^{n}\log p(C_j) + \sum_{k=1}^{K}\log p(T_k|T_k^{'}) \\
=& -\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{K} \frac{\boldsymbol{I}_{i,j}^{k}(R_{i,j}^{k} - \langle S_i, C_j, T_k\rangle)^2}{2\alpha^{-1}} + \frac{tnR\ \log\alpha}{2} \\
&- \sum_{i=1}^{m}\frac{\|S_i\|^2}{2\sigma_S^2} - m\ \log\sigma_S - \sum_{j=1}^{n}\frac{\|C_i\|^2}{2\sigma_C^2} - n\ \log\sigma_C \\
&- \sum_{k=1}^{K}\frac{\|T_k - T_k^{'}\|^2}{2\sigma_{dT}^2} - K\ \log\sigma_{dT} + \phi
\end{aligned} \tag{20}$$

where $\boldsymbol{I}_{i,j}^{k} = 1$ if $R_{i,j}^{k}$ is available and zero otherwise, tnR is the total number of ratings and ϕ is a constant. By fixing $\alpha, \sigma_S, \sigma_C$ and σ_{dT} and maximizing the log-posterior with respect to S, C, T, this objective function is equivalent to using SGD in Eq. 14 [15].

Second, the impact of using different vector lengths to encode/embed the evaluation network on the recommendation results should also be further considered. As shown in Figure 5, we test the rating prediction results with embedding vectors of five different lengths, which are 25, 50, 75, 100 and 125. The result suggests the variation of the prediction error is generally stable, and the optimal performance can be derived when the length is set to 50.

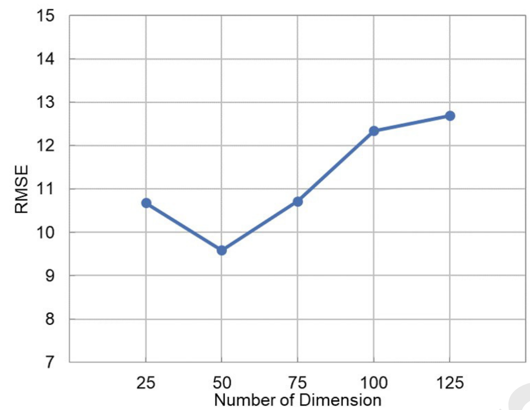


Figure 5: RMSE of our rating prediction model using embedding vectors of different lengths.

Third, our experiment results show that the proposed method has made progress in recommending new courses based on the teaching evaluation data compared to the other four models. However, we also observed some specific results of these baseline models. For example, the performance of DSE is overall worse than RTTF which is contrary to the previous report [16]. We speculate that the reason for this phenomenon is that the evaluation data characteristics used in this experiment are different from the original application scenarios. By comparing the two data sets, we find that the number of Chinese evaluation texts used in this paper is significantly smaller than the English Amazon product comment used in the original experiment (Figure 6(a)). Therefore, the method of using the teaching evaluation network for such a situation can obviously obtain more information than using only the comment text. Similarly, since Chinese course names are significantly shorter than English course names (Figure 6-b), the use of only the words in the course name to enhance the implicit habits of SVD++ is somewhat reduced. In summary, the use of common structural features in different cultural backgrounds and language habits in the field of education may be a way to achieve the better generalization of a model.

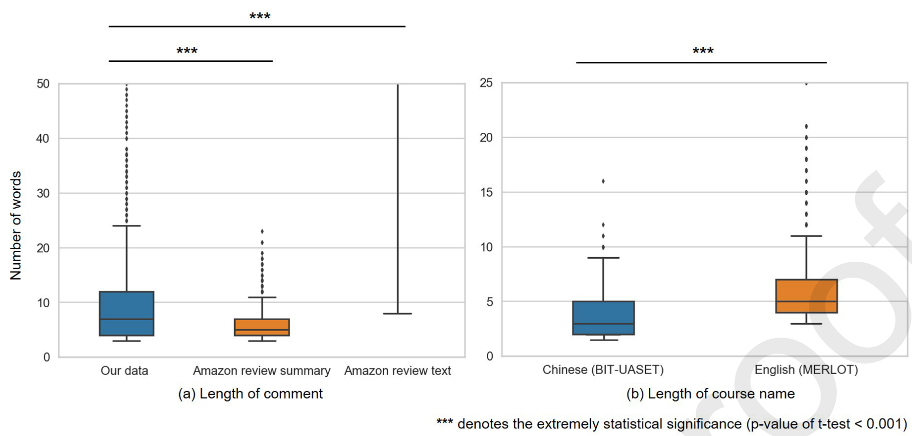


Figure 6: Difference of expression in (a) length of comment between Chinese teaching evaluation data and Amazon product comment data [65] (The main part of the "Amazon review text" is beyond the scope of the diagram due to the large difference in its value); (b) length of course name between Chinese BIT-UASET data and English MERLOT data.

6. Conclusion

In this paper, we propose a course recommendation model based on heterogeneous teaching evaluation data. We first construct a teaching evaluation network to embed the students' degree of personalization into vectors, and then we cluster the students according to these vectors. By using a BPTF based rating prediction model for course recommendation, our proposed method shows that it is practical compared with the other models on a real-world teaching evaluation dataset. Future work will include the following directions: First, this study focuses on profiling students but profiling work on courses and instructors should be further studied; Second, more behavior and attributes can be added to achieve a more accurate description of the selves of students, thereby providing a more available basis for recommending models. Third, considering the most unlabeled or semi-labeled features and strong sequential patterns in the discussed application scenario, the use of semi-supervised and reinforced learning systems may provide a more accurate application [66]. Finally, as the current method relies on offline training and recommendation, a simplified computation method is also needed in future investigation, or an optimization is required to reduce the time span during training and predicting [67].

Acknowledgements

This work was supported in part by the National Key R&D Program of China under Grant 2019YFB1406302, National Natural Science Foundation of China under Grant 61370137, 61902016, the Ministry of Education-China Mobile Research Foundation Project under Grant 2016/2-7, the Postgraduate Education Research Project of Beijing Institute of Technology under Grant 2017JYYJG-004 and supported in part by China's National Strategic Basic Research Program (973 Program) under Grant 2012CB720700.

References

[1] P. Symeonidis, D. Malakoudis, Multi-modal matrix factorization with side information for recommending massive open online courses, Expert Systems With Applications 118 (2019) 261–271.

[2] S. Wan, Z. Niu, A hybrid e-learning recommendation approach based on learners' influence propagation, IEEE Transactions on Knowledge and Data Engineering 32 (2020) 827-840.

[3] G. George, A. M. Lal, Review of ontology-based recommender systems in e-learning, Computers & Education (2019) 103642.

[4] J. K. Tarus, Z. Niu, G. Mustafa, Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning, Artificial Intelligence Review 50 (1) (2018) 21–48. doi:10.1007/s10462-017-9539-5.

[5] W. Hoiles, M. van der Schaar, Bounded off-policy evaluation with missing data for course recommendation and curriculum design, in: Proceedings of the 33nd International Conference on Machine Learning, ICML, 2016, pp. 1596–1604.

[6] X. Jing, J. Tang, Guess you like: course recommendation in moocs, in: Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017, 2017, pp. 783–789. doi:10.1145/3106426.3106478.

[7] J. Zhang, B. Hao, B. Chen, C. Li, H. Chen, J. Sun, Hierarchical reinforcement learning for course recommendation in moocs, in: The Thirty-Third AAAI Conference on Artificial Intelligence, 2019, pp. 435–442.

[8] R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, X.-Z. Gao, V. Indragandhi, A hybrid quantum-induced swarm intelligence clustering for the urban trip recommendation in smart city, Future Generation Computer Systems 83 (2018) 653–673.

[9] D. Wang, Robust policy learning control of nonlinear plants with case studies for a power system application, IEEE Transactions on Industrial Informatics 16 (3) (2020) 1733–1741. doi:10.1109/TII.2019.2925632.

[10] Y. Zhu, D. Zhao, X. Li, D. Wang, Control-limited adaptive dynamic programming for multi-battery energy storage systems, IEEE Transactions on Smart Grid 10 (4) (2018) 4235–4244. doi:10.1109/TSG.2018.2854300.

[11] B. Ma, Design of an elective course recommendation system for university environment, in: Proceedings of the 12th International Conference on Educational Data Mining, 2019.

[12] H. Liu, Y. Wang, Q. Peng, F. Wu, L. Gan, L. Pan, P. Jiao, Hybrid neural recommendation with joint deep representation learning of ratings and reviews, Neurocomputing 374 (2020) 77–85. doi:10.1016/j.neucom.2019.09.052.

[13] Q. Lin, Y. Zhu, S. Zhang, P. Shi, Q. Guo, Z. Niu, Lexical based automated teaching evaluation via students' short reviews, Computer Applications in Engineering Education 27 (1) (2019) 194–205.

[14] J. Li, C. Chang, Z. Yang, H. Fu, Y. Tang, Probability matrix factorization algorithm for course recommendation system fusing the influence of nearest neighbor users based on cloud model, in: 4th International Conference on Human Centered Computing, 2018, pp. 488–496. doi:10.1007/978-3-030-15127-0_49.

[15] J. Chambua, Z. Niu, A. Yousif, J. T. Mbelwa, Tensor factorization method based on review text semantic similarity for rating prediction, Expert Systems With Applications 114 (2018) 629–638.

[16] J. Chambua, Z. Niu, Y. Zhu, User preferences prediction approach based on embedded deep summaries, Expert Systems with Applications 132 (2019) 87–98.

[17] M. Polato, F. Aiolli, Exploiting sparsity to build efficient kernel based collaborative filtering for top-n item recommendation, Neurocomputing 268 (2017) 17–26. doi:10.1016/j.neucom.2016.12.090.

[18] A. Krstova, B. Stevanoski, M. Mihova, V. V. Ajanovski, Initialization of matrix factorization methods for university course recommendations using simrank similarities, in: 10th International Conference on ICT Innovations, 2018, pp. 172–184. doi:10.1007/978-3-030-00825-3_15.

[19] Y. Gong, Q. Zhang, X. Huang, Hashtag recommendation for multimodal microblog posts, Neurocomputing 272 (2018) 170–177. doi:10.1016/j.neucom.2017.06.056.

[20] R. Farzan, P. Brusilovsky, Social navigation support in a course recommendation system, in: V. P. Wade, H. Ashman, B. Smyth (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 91–100. doi:10.1007/11768012_11.

[21] K. H. Tsai, T. Hsieh, T. K. Chiu, M. Lee, T. I. Wang, Automated course composition and recommendation based on a learner intention, in: Pro-

ceedings of the 7th IEEE International Conference on Advanced Learning Technologies, 2007, pp. 274–278. doi:10.1109/ICALT.2007.80.

[22] A. Pongpech, M. E. Orlowska, S. W. Sadiq, Personalized courses recommendation functionality for flex-el, in: Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies, 2007, pp. 631–633. doi:10.1109/ICALT.2007.203.

[23] F. Durao, M. P. Caraciolo, B. Melo, S. R. d. L. Meira, Towards effective course-based recommendations for public tenders, International Journal of Knowledge and Web Intelligence 4 (2-3) (2013) 260–278.

[24] A. Elbadrawy, G. Karypis, Domain-aware grade prediction and top-n course recommendation, in: Proceedings of the 10th ACM Conference on Recommender Systems, 2016, pp. 183–190. doi:10.1145/2959100.2959133.

[25] H. Thanh-Nhan, H. Nguyen, N. Thai-Nghe, Methods for building course recommendation systems, in: 2016 Eighth International Conference on Knowledge and Systems Engineering, 2016, pp. 163–168. doi:10.1109/KSE.2016.7758047.

[26] E. S. Khorasani, Z. Zhenge, J. Champaign, A markov chain collaborative filtering model for course enrollment recommendations, in: 2016 IEEE International Conference on Big Data, 2016, pp. 3484–3490. doi:10.1109/BigData.2016.7841011.

[27] E. L. Lee, T. Kuo, S. Lin, A collaborative filtering-based two stage model with item dependency for course recommendation, in: 2017 IEEE International Conference on Data Science and Advanced Analytics, 2017, pp. 496–503. doi:10.1109/DSAA.2017.18.

[28] J. Li, C. Chang, Z. Yang, H. Fu, Y. Tang, Probability matrix factorization algorithm for course recommendation system fusing the influence of nearest neighbor users based on cloud model, in: 4th Interna-

tional Conference on Human Centered Computing, 2018, pp. 488–496. doi:10.1007/978-3-030-15127-0_49.

[29] L. Huang, C. Wang, H. Chao, J. Lai, P. S. Yu, A score prediction approach for optional course recommendation via cross-user-domain collaborative filtering, IEEE Access 7 (2019) 19550–19563. doi:10.1109/ACCESS.2019.2897979.

[30] J. Lin, Y. Li, J. Lian, A novel recommendation system via l0-regularized convex optimization, Neural Computing and Applications 32 (6) (2020) 1–15. doi:10.1007/s00521-019-04213-w.

[31] R. G. Apaza, E. V. Cervantes, L. C. Quispe, J. O. Luna, Online courses recommendation based on LDA, in: Proceedings of the 1st Symposium on Information Management and Big Data, 2014, pp. 42–48.

[32] Y. Dai, Y. Asano, M. Yoshikawa, Course content analysis: An initiative step toward learning object recommendation systems for MOOC learners, in: Proceedings of the 9th International Conference on Educational Data Mining, 2016, pp. 347–352.

[33] C. Huang, R. Chen, L. Chen, Course-recommendation system based on ontology, in: International Conference on Machine Learning and Cybernetics, 2013, pp. 1168–1173. doi:10.1109/ICMLC.2013.6890767.

[34] J. Antony, S. Ramakrishnan, M. T. Joy, R. Thottupuram, Sms based e-assessments enabling better student engagement, evaluation and recommendation services in e-learning making use of fuzzy rules and course ontologies, in: 2012 IEEE Fourth International Conference on Technology for Education, 2012, pp. 119–125. doi:10.1109/T4E.2012.9.

[35] M. E. Ibrahim, Y. Yang, D. Ndzi, Using ontology for personalised course recommendation applications, in: 17th International Conference on Computational Science and Its Applications, 2017, pp. 426–438. doi:10.1007/978-3-319-62392-4_31.

[36] M. E. Ibrahim, Y. Yang, D. L. Ndzi, G. Yang, M. Al-Maliki, Ontology-based personalized course recommendation framework, IEEE Access 7 (2019) 5180–5199. doi:10.1109/ACCESS.2018.2889635.

[37] S. B. Aher, L. Lobo, Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data, Knowledge-Based Systems 51 (2013) 1–14.

[38] J. Xu, T. Xing, M. Van Der Schaar, Personalized course sequence recommendations, IEEE Transactions on Signal Processing 64 (20) (2016) 5340–5352.

[39] H. Zhang, T. Huang, Z. Lv, S. Liu, Z. Zhou, Mcrs: A course recommendation system for moocs, Multimedia Tools and Applications 77 (6) (2018) 7051–7069.

[40] Y. Hou, P. Zhou, J. Xu, D. O. Wu, Course recommendation of MOOC with big data support: A contextual online learning approach, in: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops, 2018, pp. 106–111. doi:10.1109/INFCOMW.2018.8406936.

[41] A. Polyzou, A. N. Nikolakopoulos, G. Karypis, Scholars walk: A markov chain framework for course recommendation, in: Proceedings of the 12th International Conference on Educational Data Mining, 2019.

[42] A. Parameswaran, P. Venetis, H. Garcia-Molina, Recommendation systems with complex constraints: A course recommendation perspective, ACM Transactions on Information Systems 29 (4) (2011) 20.

[43] S. Yang, Y. Feng, L. Zou, A. Jia, D. Zhao, Taxonomy induction and taxonomy-based recommendations for online courses, in: Proceedings of the 15th ACM/IEEE-CE Joint Conference on Digital Libraries, 2015, pp. 267–268. doi:10.1145/2756406.2756968.

[44] E. García, C. Romero, S. Ventura, C. De Castro, An architecture for making recommendations to courseware authors using association rule mining and

collaborative filtering, User Modeling and User-Adapted Interaction 19 (1-2) (2009) 99–132.

[45] P. Chang, C. Lin, M. Chen, A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems, Algorithms 9 (3) (2016) 47. doi:10.3390/a9030047.

[46] Y. Ng, J. Linn, Crsrecs: A personalized course recommendation system for college students, in: 8th International Conference on Information, Intelligence, Systems & Applications, 2017, pp. 1–6. doi:10.1109/IISA.2017.8316368.

[47] J. Lin, H. Pu, Y. Li, J. Lian, Sparse linear method based top-n course recommendation system with expert knowledge and l_0 regularization, in: Third International Conference on Human Centered Computing, 2017, pp. 130–138. doi:10.1007/978-3-319-74521-3_15.

[48] A. Esteban, A. Zafra, C. Romero, Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization, Knowledge-Based Systems 194 (2020) 105385. doi:https://doi.org/10.1016/j.knosys.2019.105385.

[49] S. Wan, Z. Niu, A learner oriented learning recommendation approach based on mixed concept mapping and immune algorithm, Knowlege-Based Systems 103 (2016) 28–40. doi:10.1016/j.knosys.2016.03.022.

[50] J. K. Tarus, Z. Niu, A. Yousif, A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining, Future Generation Computer Systems 72 (2017) 37–48. doi:10.1016/j.future.2017.02.049.

[51] S. Wan, Z. Niu, An e-learning recommendation approach based on the self-organization of learning resource, Knowledge-Based Systems 160 (2018) 71–87. doi:10.1016/j.knosys.2018.06.014.

[52] G. Wisker, The good supervisor: supervising postgraduate and undergraduate research for doctoral theses and dissertations, R & D Management 36 (2) (2012) 220–221.

[53] T. Mikolov, K. Chen, G. S. Corrado, J. Dean, Efficient estimation of word representations in vector space, in: International Conference on Learning Representations, 2013.

[54] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014, pp. 701–710.

[55] A. Grover, J. Leskovec, node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 855–864.

[56] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Advances in Neural Information Processing Systems, 2013, pp. 3111–3119.

[57] M. Ester, H. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, 1996, pp. 226–231.

[58] A. Mnih, R. R. Salakhutdinov, Probabilistic matrix factorization, in: Advances in Neural Information Processing Systems, 2008, pp. 1257–1264.

[59] J. Liu, C. Wu, W. Liu, Bayesian probabilistic matrix factorization with social relations and item contents for recommendation, Decision Support Systems 55 (3) (2013) 838–850.

[60] R. Salakhutdinov, Bayesian probabilistic matrix factorization using markov chain monte carlo, in: International Conference on Machine Learning, 2008.

[61] R. Salakhutdinov, A. Mnih, Probabilistic matrix factorization, in: International Conference on Neural Information Processing Systems, 2007.

[62] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, Siam Review 51 (3) (2009) 455–500.

[63] J. Kossaifi, Y. Panagakis, A. Anandkumar, M. Pantic, Tensorly: Tensor learning in python, Journal of Machine Learning Research 20 (26) (2019) 1–6.

[64] Y. Koren, Factorization meets the neighborhood: A multifaceted collaborative filtering model, in: ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2008.

[65] R. He, J. McAuley, Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering, in: Proceedings of the 25th International Conference on World Wide Web, 2016, pp. 507–517.

[66] H. Li, D. Liu, D. Wang, Manifold regularized reinforcement learning, IEEE Transactions on Neural Networks and Learning Systems 29 (4) (2017) 932–943. doi:10.1109/TNNLS.2017.2650943.

[67] D. Wang, M. Ha, J. Qiao, Self-learning optimal regulation for discrete-time nonlinear systems under event-driven formulation, IEEE Transactions on Automatic Control (2019) (Early Access).doi:10.1109/TAC.2019.2926167.