

Sparse Linear Method Based Top-N Course Recommendation System with Expert Knowledge and L_0 Regularization

Jinjiao Lin^{1,2}, Haitao Pu³, Yibin Li¹, and Jian Lian³(✉)

¹ School of Control and Engineering, Shandong University, Jinan, China

² Shandong University of Finance and Economics, Jinan, China

³ Shandong University of Science and Technology, Qingdao, China
lianjianlianjian@163.com

Abstract. In this paper, we propose an approach of course recommender system for the subject of information management speciality in China. We collect the data relative to the course enrollment for specific set of students. The sparse linear method (SLIM) is introduced in our approach to generate the top-N recommendations of courses for students. Furthermore, the L_0 regularization terms were presented in our proposed optimization method based on the observation of the entries in recommendation system matrix. Expert knowledge based comparing experiments between state-of-the-art methods and our method are conducted to evaluate the performance of our method. Experimental results show that our proposed method outperforms state-of-the-art methods both in accuracy and efficiency.

Keywords: Course recommender system · Sparse linear method
Expert knowledge

1 Introduction

The emergence and rapid development of Internet have greatly affected the traditional viewpoint on choosing courses by providing detailed course information. As the number of courses conforming to the students' has tremendously increased, the above-mentioned problem has become how to determine the courses mostly suitable for the students accurately and efficiently. A plethora of methods and algorithms [2, 3, 11, 15] for course recommendation have been proposed to deal with this problem. Most of the methods designed for recommendation system can be grouped into three categories, including collaborative [1, 8], content-based [7, 14], and knowledge-based [5, 8, 17], which have been applied in different fields such as [4] proposed a collaborative filtering embedded with an artificial immune system to the course recommendation for college students. The rating from professor was exploited as ground truth to examine the results.

Inspired by the idea form [4] and the optimization framework in [9], we propose a sparse linear based method for top-N course recommendation with expert knowledge as the ground truth. This method extracts the coefficient matrix for the courses in the

recommender system from the student/course matrix by solving a regularized optimization problem. The sparseness is exploited to represent the sparse characteristics of recommendation coefficient matrix. Sparse linear method (SLIM) [9] was proposed to top-N recommender systems, which is rarely exploited in course recommender systems. Due to the characteristics of course recommendation system in Chinese University, our method focuses on the accuracy more than the efficiency. It is different from the previously proposed SLIM based methods [6, 9, 10, 18], which mainly addresses the real-time applications of top-N recommender systems. The framework of our proposed course recommender system is shown in Fig. 1.

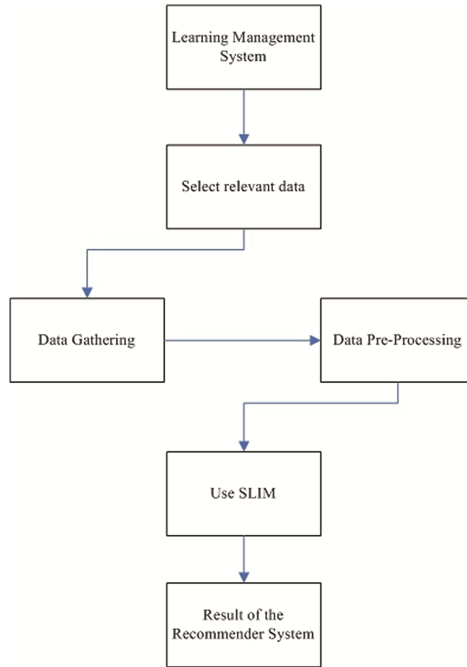


Fig. 1. The framework of our proposed course recommender system

According to our observation about common recommendation system matrix, most of the entries are assigned the same value (zero or one), and the gradients of neighboring entries also hold the same value (zero or one). Therefore, the sparse counting strategy of L_0 regularization terms [16] were included into the optimization framework of SLIM. The L_0 terms can globally constrain the non-zero values of entries and the gradients in the recommendation system matrix, which is the main contribution of our proposed method. Different from the previously proposed regularization terms (the L_1 and L_2 terms), the L_0 term can maintain the subtle relationship between the entries in recommendation system matrix.

After the process of data gathering as shown in Fig. 1, comparing experiments between state-of-the-art methods and our method are conducted. Consequently, both the

experimental results of state-of-the-art methods and our method are evaluated with the course recommendations presented by seven experts with voting strategy.

The rest of the paper is organized as follows. In Sect. 2, we describe the details of our proposed method. In Sect. 3 the dataset that we used in our experiments and the experimental results are presented. In Sect. 4 the discussion and conclusion are given.

2 Our Method

2.1 The Formation of the Method

In the following content, t_j and s_i are introduced to denote each course and each student in course recommender system, respectively. The whole student-course taken will be represented by a matrix A of size $m \times n$, in which the entry is 1 or 0 (1 denotes that the student has taken the course, 0 vice versa).

In this paper, we introduce a Sparse Linear Method (SLIM) to implement top-N course recommendation. In this approach, the score of course recommendation on each un-taken student/course item t_j of a student s_i is computed as a sparse aggregation of items that have been taken by s_i , which is shown in Eq. (1).

$$\bar{a}_{ij} = a_i^T w_j \quad (1)$$

where \bar{a} is the initial course selection of a specific student and w_j is the sparse vector of aggregation coefficients. The model of SLIM with matrix is represented as:

$$\bar{A} = AW \quad (2)$$

Where \overline{A} is the initial value of student/course matrix, A denotes the latent binary student-course item matrix, W denotes the $n \times n$ sparse matrix of aggregation coefficients, in which j -th column corresponds to w_j as in Eq. (1), and each row of $C(c_i)$ is the course recommendation scores on all courses for student s_i . The final course recommendation result of each student is completed through sorting the non-taken courses in decreasing order, and the top-N courses in the sequences are recommended.

In our method, the initial student/course matrix is extracted from the learning management system of a specific University in China. With the extracted student/course matrix of size $m \times n$, the sparse matrix W size of $n \times n$ in Eq. (2) is iteratively optimized by alternate minimization method. Different from the objective function previously proposed in [9] shown in Eq. (3), our proposed method is shown in Eq. (4).

$$\min_{w \geq 0} \frac{1}{2} \|A - AW\|_2^2 + \frac{\beta_1}{2} \|W\|_F^2 + \lambda_1 \|A\|_1 \quad (3)$$

$$\min_{w \geq 0} \frac{1}{2} \|A - AW\|_2^2 + \frac{\beta_2}{2} \|W\|_F^2 + \lambda_2 \|A\|_0 + \mu |\nabla A|_0 \quad (4)$$

Where $\|\cdot\|_F$ denotes the Frobenius norm for matrix, $\|W\|_1$ is the item-wise L_1 norm, $\|W\|_0$ denotes the entry-wise L_0 norm that stands for the number of entries with zero

value. The data term $\|A - AW\|$ is exploited to measure the difference between the calculated model and the training dataset. The L_F -norm, L_1 -norm, and L_0 -norm are exploited to regularize the entries of the coefficient matrix W , A , and ∇A , respectively. The parameters β_1 , β_2 , λ_2 , and μ are used to constrain the weights of regularization terms in the objective functions.

In our proposed final objective function, the L_F norm is introduced to transfer the optimization problem into elastic net problem [19], which prevents the potential over fitting. Moreover, the L_1 norm in Eq. (3) is changed to L_0 norm in our proposed objective function. This novel norm L_0 [12, 13, 16] is introduced to constrain the sparseness of the A and ∇A .

Due to the independency of the columns in matrix W , the final objective function in Eq. (4) is decoupled into a set of objective functions as follows:

$$\min_{w_j \geq 0, w_{j,j}=0} \frac{1}{2} \|a_j - a_j w_j\|_2^2 + \frac{\beta_2}{2} \|w_j\|_F^2 + \lambda_2 \|a_j\|_0 + \mu \|\nabla a_j\|_0 \tag{5}$$

where a_j is the j -th column of matrix A , w_j denotes j -th column of matrix W . As there are two unknown variables in each Eq. (5), which is a typical ill-posed problem. Thus, this problem need to be solved by alternate minimization method. In each iteration, one of the two variables is fixed and the other variable is optimized.

2.2 The Solver of Our Proposed Method

Subproblem1: computing w_j

The w_j computation sub-problem is represented by the minimization of Eq. (6):

$$\frac{1}{2} \|a_j - a_j w_j\|_2^2 + \frac{\beta_2}{2} \|w_j\|_F^2 \tag{6}$$

Through eliminating the L_0 terms in Eq. (5), the function Eq. (6) has a global minimum, which can be computed by gradient descent. The analytical solution to Eq. (6) is shown in Eq. (7):

$$w_j = F^{-1} \left(\frac{F(a_j)}{F(a_j) + \frac{\beta_2}{2} (F(\partial_x)^* \cdot F(\partial_x) + F(\partial_y)^* \cdot F(\partial_y))} \right) \tag{7}$$

where $F(\cdot)$ and $F^{-1}(\cdot)$ denotes the Fast Fourier Transform (FFT) and reverse FFT, respectively. $F(\cdot)^*$ is the complex conjugate of $F(\cdot)$.

Sub-problem 2: computing a_j and ∇a_j

With the intermediate outcome of w_j , the a_j and ∇a_j can be computed by Eq. (8):

$$\frac{1}{2}\|a_j - a_j w_j\|_2^2 + \lambda_2 \|a_j\|_0 + \mu \|\nabla a_j\|_0 \quad (8)$$

By introducing two auxiliary variables h and v corresponding to the column vector a_j and ∇a_j . The sub-problem can be transformed into Eq. (9):

$$\frac{1}{2}\|a_j - a_j w_j\|_2^2 + \lambda_2 \|a_j - h\|_2^2 + \mu \|\nabla a_j - v\|_2^2 + \lambda (\|h\|_0 + \|v\|_0) \quad (9)$$

To testify the performance of our proposed method, comparing experiments between state-of-the-art methods and our method are carried out with gathered dataset and expert knowledge. In the following section, the experiments are described in detail.

3 Experimental Results

3.1 Datasets

In order to testify the performance of our proposed method and implement the method in practical scenarios, we gather the data from five classes of information management specialty for the learning management system of our University. The data records of the

Table 1. The initial dataset from the five classes

No.	SPSS	CH	Eng	LA	PT	DB	CC	PE	C	Acc	CS
1	1	1	1	0	1	0	1	1	0	1	0
2	0	1	1	0	1	0	1	1	1	1	0
3	0	1	1	0	1	1	0	1	0	1	0
4	0	1	1	0	1	1	0	1	1	1	0
5	1	1	1	1	1	0	0	1	1	1	0
6	1	1	1	1	1	0	0	1	1	1	1
7	0	1	1	1	1	0	0	1	1	1	1
8	1	1	1	1	1	1	0	1	0	1	1
9	1	1	1	0	1	1	1	1	0	1	0
10	1	1	1	0	1	1	1	1	0	1	0
11	0	1	1	0	1	0	1	1	0	1	0
12	0	1	1	0	1	1	0	1	0	1	1
13	0	1	1	1	1	0	1	1	0	1	1
14	1	1	1	1	1	1	1	1	0	1	1
15	1	1	1	1	1	1	1	1	0	1	1
16	0	1	1	0	1	1	1	1	0	1	1
17	1	1	1	1	1	1	1	1	0	1	1
18	0	1	1	1	1	0	1	1	0	1	0
19	1	1	1	0	1	1	1	1	0	1	0
20	0	1	1	1	1	0	1	1	0	1	1

courses and students were extracted from the Department of Management Information System, Shandong University of Finance and Economics and the Department of Electronic Engineering Information Technology at Shandong University of Sci&Tech. The most important information of the courses and students is mainly about the grades corresponding to the courses. All of the students from the information management specialty are freshmen in our University. Most of them have taken the courses of the first year in their curriculum except three students have failed to go up to the next grade. Thus, firstly we eliminate the records of the three students. Meanwhile, we collect the knowledge including the programming skill that they have mastered through a questionnaire. The courses that they have taken and the content that have grasped are combined in the final dataset. A part of the dataset is shown in Table 1, where 1 denotes that the s_i student has mastered the t_j course, and 0 denotes the opposite.

After gathering the data of the students from the five classes, comparing experiments between state-of-the-art methods and our method are conducted. We choose several state-of-the-art methods including collaborative filtering methods itemkNN, userkNN, and the matrix factorization methods PureSVD.

3.2 Measurement

The knowledge from several experts on the courses in information management specialty are adopted as ground truth in the experimental process. To measure the performance of the comparing methods, we introduce the Hit Rate (HR) and the Average Reciprocal Hit-Rank (ARHR) in the experiments, which are defined as shown in Eqs. (11) and (12).

$$HR = \frac{\#hits}{\#students} \quad (11)$$

where $\#hits$ denotes the number of students whose course in the testing set is recommended by the expert, too. $\#students$ denotes the number of all students in the dataset.

$$ARHR = \frac{1}{\#students} \sum_{i=1}^{\#hits} \frac{1}{p_i} \quad (12)$$

Where p_i is the ordered recommendation list.

3.3 Experimental Results

In this section, the experimental results calculated from the practical dataset. Table 2 shows the experimental results of the comparing methods in top-N course recommendation.

Table 2. The performance of the comparing methods

Methods	HR_1	$ARHR_1$	HR_2	$ARHR_2$	HR_3	$ARHR_3$	HR_4	$ARHR_4$	HR_5	$ARHR_5$
itemkN	0.18	0.13	0.19	0.14	0.20	0.13	0.18	0.13	0.19	0.14
N										
Itemprob	0.21	0.15	0.19	0.16	0.21	0.14	0.19	0.12	0.17	0.13
PureSVD	0.09	0.11	0.10	0.12	0.12	0.12	0.17	0.14	0.18	0.15
SLIM	0.24	0.16	0.17	0.18	0.24	0.19	0.16	0.14	0.17	0.15
ours	0.27	0.17	0.19	0.17	0.25	0.18	0.20	0.14	0.19	0.15

Where $HR_i, ARHR_i$ denotes the performance for $class_i$, respectively. The experimental results shown in Table 2 demonstrate that our proposed method outperforms state-of-the-art methods in most of course recommendations both in the HR and ARHR. It shows that the sparse regularization term based on the prior knowledge from the observation in our method are suitable for solving the problem of course recommendation.

In order to illustrate the performance of our proposed method according to the number of courses and topics included in the experimental testing. It shows in Fig. 2 that a higher accuracy is obtained when the number of courses increases. Meanwhile, the courses included in our experiments are divided into 32 different topics, Fig. 3 shows that the accuracy is also higher when there are more relative courses.

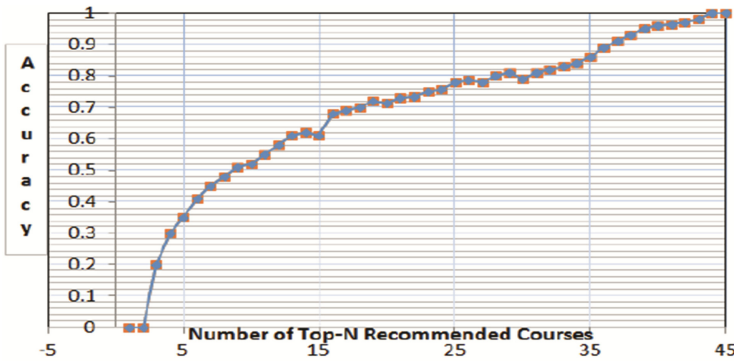


Fig. 2. Accuracy of our proposed recommendation system method due to the number of courses

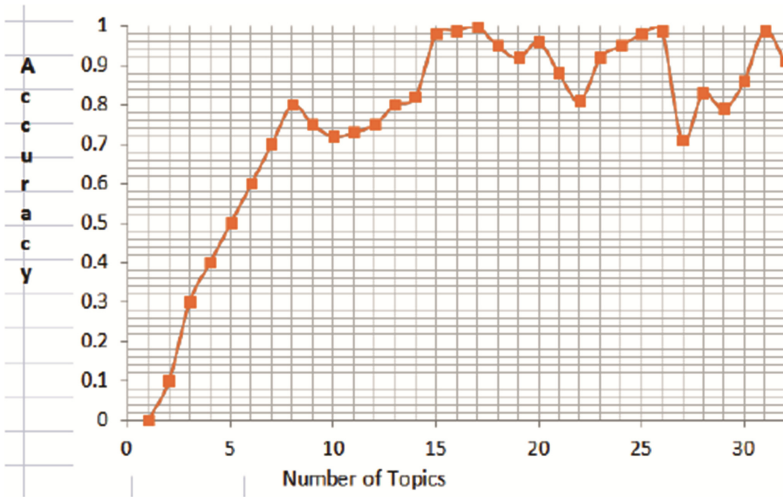


Fig. 3. Accuracy of our proposed recommendation system method due to the number of topics

4 Conclusion

In this paper, we propose an approach of course recommendation. In our method, the SLIM was introduced and a novel L_0 regularization term was exploited in SLIM. Meanwhile, the alternate minimization strategy is exploited to optimize the outcome of our method. To testify the performance of our method, comparing experiments on students from five different classes between state-of-the-art methods and our method are conducted. The experimental results show that our method outperforms the other previously proposed methods.

The proposed method was be mainly used to implement the course recommendation for the Universities in China. However, it also can b exploited in other relative fields. In the future, more applications of our approach would be investigated. Other future work includes the modification of the objective function in our method including the other regularization terms and different optimization strategy.

Acknowledgments. This work was financially supported by the Teaching Reform Research Project of Undergraduate Colleges and Universities of Shandong Province (2015M111, 2015M110, Z2016Z036) and the Teaching Reform Research Project of Shandong University of Finance and Economics (2891470), Teaching Reform Research Project of Undergraduate Colleges and Universities of Shandong Province (2015M136), SDUST Young Teachers Teaching Talent Training Plan (BJRC20160509); SDUST Excellent Teaching Team Construction Plan; Teaching research project of Shandong University of Science and Technology (JG201509 and qx2013286); Shandong Province Science and Technology Major Project (No. 2015ZDX X0801A02).

Disclosures. The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Ahn, H.: Utilizing popularity characteristics for product recommendation. *Int. J. Electron. Commer.* **11**(2), 59–80 (2006)
2. Albadaerah, A., Alsakran, J.: An automated recommender system for course selection. *Int. J. Adv. Comput. Sci. Appl.* **7**(3) (2016)
3. Aher, S.B., Lobo, L.M.R.J.: A comparative study of association rule algorithms for course recommender system in e-learning. *Int. J. Comput. Appl.* **39**(1), 48–52 (2012)
4. Chang, P., Lin, C., Chen, M.: A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems. *Algorithms* **9**(3), 47 (2016)
5. Chen, Y., Cheng, L., Chuang, C.: A group recommendation system with consideration of interactions among group members. *Expert Syst. Appl.* **34**(3), 2082–2090 (2008)
6. Christakopoulou, E., Karypis, G.: HOSLIM: higher-order sparse linear method for top- N recommender systems. In: Tseng, V.S., Ho, T.B., Zhou, Z.-H., Chen, A.L.P., Kao, H.-Y. (eds.) PAKDD 2014. LNCS (LNAI), vol. 8444, pp. 38–49. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-06605-9_4
7. Kim, E., Kim, M., Ryu, J.: Collaborative filtering based on neural networks using similarity. In: Wang, J., Liao, X.-F., Yi, Z. (eds.) ISNN 2005. LNCS, vol. 3498, pp. 355–360. Springer, Heidelberg (2005). https://doi.org/10.1007/11427469_57
8. McLaughlin, M., Herlocker, J.L.: A collaborative filtering algorithm and evaluation metric that accurately model the user experience, pp. 329–336 (2004)
9. Ning, X., Karypis, G.: SLIM: sparse linear methods for top- n recommender systems, pp. 497–506 (2011)
10. Ning, X., Karypis, G.: Sparse linear methods with side information for top- n recommendations, pp. 155–162 (2012)
11. Omahony, M.P., Smyth, B.: A recommender system for on-line course enrolment: an initial study, pp. 133–136 (2007)
12. Pan, J., Hu, Z., Su, Z., Yang, M.: Deblurring text images via L₀-regularized intensity and gradient prior, pp. 2901–2908 (2014)
13. Pan, J., Lim, J., Su, Z., Yang, M.H.: L₀-regularized object representation for visual tracking. In: Proceedings of the British Machine Vision Conference. BMVA Press (2014)
14. Philip, S., Shola, P.B., John, A.O.: Application of content-based approach in research paper recommendation system for a digital library. *Int. J. Adv. Comput. Sci. Appl.* **5**(10) (2014)
15. Sunil, L., Saini, D.K.: Design of a recommender system for web based learning, pp. 363–368 (2013)
16. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via L₀ gradient minimization. In: International Conference on Computer Graphics and Interactive Techniques, vol. 30, No. 6 (2011). 174
17. Yap, G., Tan, A., Pang, H.: Dynamically-optimized context in recommender systems, pp. 265–272 (2005)
18. Zheng, Y., Mobasher, B., Burke, R.: CSLIM: contextual slim recommendation algorithms, pp. 301–304 (2014)
19. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc. Ser. B-Stat. Methodol.* **67**(2), 301–332 (2005)