Personalized Multi-relational Matrix Factorization Model for Predicting Student Performance

Prema Nedungadi and T.K. Smruthy

Abstract Matrix factorization is the most popular approach to solving prediction problems. However, in the recent years multiple relationships amongst the entities have been exploited in order to improvise the state-of-the-art systems leading to a multi relational matrix factorization (MRMF) model. MRMF deals with factorization of multiple relationships existing between the main entities of the target relation and their metadata. A further improvement to MRMF is the Weighted Multi Relational Matrix Factorization (WMRMF) which treats the main relation for the prediction with more importance than the other relations. In this paper, we propose to enhance the prediction accuracy of the existing models by personalizing it based on student knowledge and task difficulty. We enhance the WMRMF model by incorporating the student and task bias for prediction in multi-relational models. Empirically we have shown using over five hundred thousand records from Knowledge Discovery dataset provided by Data Mining and Knowledge Discovery competition that the proposed approach attains a much higher accuracy and lower error(Root Mean Square Error and Mean Absolute Error) compared to the existing models.

1 Introduction

Predicting student performance plays an important role in helping students learn better [7]. By this prediction, teachers may better understand the student knowledge in each domain and thus help the students in deep learning. The problem of predicting performance of the students has been efficiently solved using different matrix

P. Nedungadi
Amrita CREATE, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India
e-mail: prema@amrita.edu

T.K. Smruthy(✉)
Dept. of Computer Science, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India
e-mail: smruthykrishnan@gmail.com

S. Berretti et al. (eds.), *Intelligent Systems Technologies and Applications*,
Advances in Intelligent Systems and Computing 384,
DOI: 10.1007/978-3-319-23036-8_15

factorization models in the past. A simple matrix factorization deals with only one main target relation (such as Student performs task). This relation is represented by a matrix where each cell represents the association between a particular student and a task. The matrix so constructed is factorized to yield two smaller matrices each representing factors of students and tasks[14]. However such a prediction depends on many other relations as well. Thus, this results in a multi relational matrix factorization model where the factors of the target entities are determined by collectively factorizing both the target relation and the dependent relations [6]. Further, weights are assigned to each relation based on their importance in the prediction problem. This results in Weighted Multi Relational Matrix Factorization (WMRMF) [16]. In this paper, we propose a biased WMRMF model which enhances the prediction accuracy of the state-of-the-art systems. For a student performance prediction problem, there are two main bias involved namely the student bias which explains a generic probability of a student in performing tasks and the task bias which explains the degree of difficulty of the task [16]. By considering the biases for prediction in multi relational model, we improve the student performance prediction accuracy drastically. Also, the state-of-the-art systems considers sum of product of latent factors, bias and global average for performance prediction. Empirically we show that considering only the bias and neglecting the global average provides more accurate results than considering both together as done in simple matrix factorization models. Hence global average has been used only for the cold start problems while for the rest, the sum of product of latent factors and biases are used.

2 Related Work

For the student performance prediction problem, many works have been done using classification and regression modeling mainly using Bayesian networks, decision tree [2][7]. There are also many models using knowledge tracing (KT) which works to find four basic parameters namely prior knowledge, learn rate, slip and guess factors using several techniques like Expectation Minimization approach, brute-force approach. [3][8]. Later, models using matrix factorization techniques were proposed which took parameters like slip and guess factors implicitly into account. The very initial model was to consider the relation Student performs tasks as a matrix where each cell value represents the performance of a student for a particular task. The matrix so constructed was factorized into two smaller latent factor matrices and prediction was done by calculating the interaction between the factors [5]. Biased model for single matrix factorization was also developed to take into consideration the user effect and item effect [5]. Both the models prove very effective for sparse matrices but only the target relation is optimized neglecting any dependent relations. Later, prediction problems were solved using multi relational models which exploited the multiple relationships existing between entities and they prove effective in recommender systems [6][11][1]. These models were used explicitly in student performance prediction domain and further, a weighted multi relational matrix factorization model was proposed and shown to be successful in terms of prediction accuracy [14][15][16]. But none of these models considered either student bias or the task bias. Further, an enhancement to matrix factorization approach was tensor factorization [4]. Tensor factorization has been used for student performance prediction which also takes the temporal or the sequential effect into consideration [13]. One of our previous work proposes an enhanced and efficient tensor decomposition technique for student performance prediction [9]. But the relationships that can be considered while modeling as a tensor are limited when compared to exploiting multiple relationships as matrices. Also, while modeling as tensors, as the number of relationships considered increases, it leads to a high space complexity. In our paper, we implement an enhanced weighted multi relational matrix factorization model considering both the student and task bias along with the latent factors that gives promising results in terms of prediction accuracy.

3 Personalized Multi-relational Matrix Factorization Model

Let there be N entities (E1, E2.. EN) e.g. (Student, Task) and M relations (R1, R2.. RM) (performs, requires) existing in a student performance prediction problem. Our aim is to predict unobserved values between two entities from already observed values, such as in a student performance prediction problem, our aim is to predict how well a student can perform a task. For this many models have been proposed. Before we move on to multi relational models, let us review the simple matrix factorization approach. A student-performs-task matrix R is approximated as a product of two smaller matrices W_1 (student) and W_2 (task) where each row contains F latent factors describing that row. Then, w_s and w_i describes the vectors of W_1 and W_2 and their elements are denoted by w_{sf} and w_{if} . Then the performance of student s for a task i can be predicted as:

$$\hat{p} = \sum_{f=1}^{F} w_{sf} w_{if} = w_s w_i^T \tag{1}$$

where $\hat{p}$ is the predicted performance value. w_s and w_i can be learnt by optimizing the objective function :

$$O^{MF} = \sum_{(s,i)\in R} error_{si}^2 + \lambda(\| W_1 + W_2 \|_F^2) \tag{2}$$

where $\| \ . \ \|_F^2$ is the Frobenius norm and λ is the regularization term to avoid over fitting, and $error_{si}^2$ is calculated as the difference between the actual performance value and the predicted value for each student-task combination and is given as:

$$error_{si}^2 = ((R)_{si} - w_s w_i^T)^2 \tag{3}$$

where $(R)_{si}$ represents actual performance value of student s for task i, Optimization is done using stochastic gradient descent model. Let us now look at the multi relational models.

3.1 Base Model 1: Multi Relational Matrix Factorization (MRMF)

Different relationships are drawn out from the domain under consideration and the factor matrices of each entity are found by a collective matrix factorization. As discussed earlier, in a system with N entities (E1, E2..EN) and M relations (R1, R2..RM), let $W_n(n \in N)$ be the latent factor matrices of each of the entities with F latent factors. These latent factors describe the entity and are built by considering every relation that the entity is associated with [6][16]. In such a system, the model parameters are learnt using the optimization function:

$$O^{MRMF} = \sum_{r=1}^{M} \sum_{(s,i)\in R_r} ((R_r)_{si} - w_{r1s} w_{r2i}^{T})^2 + \lambda(\sum_{j=1}^{N} \parallel W_j \parallel_F^2) \tag{4}$$

3.2 Base Model 2: Weighted Multi Relational Matrix Factorization (WMRMF)

In the previous model, every relation is given equal weightage [16]. But practically, weight of the relations change for different target relations, such as in a student performance prediction, student performs task is the main relation and hence this relation should be given more weightage. Thus a weight factor (θ) is added to the previous MRMF model and the optimization function is:

$$O^{WMRMF} = \sum_{r=1}^{M} \theta_r \sum_{(s,i)\in R_r} ((R_r)_{si} - w_{r1s} w_{r2i}^{T})^2 + \lambda(\sum_{j=1}^{N} \parallel W_j \parallel_F^2) \tag{5}$$

θ is given a value 1 for the main relation and for the rest of the relation we assign a lower value.

3.3 Proposed Approach 1: Personalized Multi Relational Matrix Factorization Using Bias and Global Average

In this paper, we propose an enhancement to the weighted multi relational model, increasing the accuracy of prediction by considering bias terms and global average for prediction. The student bias explains the probability of a student in performing tasks and the task bias explains the degree of difficulty of the task. Global average (μ) explains the average performance of all students and tasks in the training set considered [13]. The two biases and global average are added along with the dot product of latent factors while predicting performance which is given as:

$$\hat{p} = \mu + b_s + b_t + \sum_{k=1}^{K} w_{sk} h_{ik} \tag{6}$$

Thus the optimization function now becomes:

$$O^{B-WMRMF} = \sum_{r=1}^{M} \theta_r \sum_{(s,i)\in R_r} ((R_r)_{si} - (\mu + b_s + b_i + w_{r1s} w_{r2i}^T))^2 + \lambda(\sum_{j=1}^{N} \| W_j \|_F^2 + b_s^2 + b_i^2) \tag{7}$$

where b_s and b_t are the student and task bias respectively. Bias terms have been introduced in the single matrix factorization models in the past [13]. But to the best of our knowledge, this is the first paper where biases are considered in a multi relational environment. Student bias is calculated as the average of deviation of performance of student s from global average and task bias is calculated as the average of deviation of performance of task i from global average.

3.4 Proposed Approach 2: Personalized Multi Relational Matrix Factorization Using Only Bias

Along with the bias terms, the global average has also been considered for performance prediction in the previous enhancement proposed (Proposed approach 1) and this model has proved to produce higher accuracy when compared to the base models. We also propose a further enhancement to proposed approach 1 by considering only bias along with the latent factor products. Empirically we have shown that neglecting the global average and considering just the bias and the latent factors lead to better prediction accuracy in multi relational factorization. Global average has been considered for cold start problems only. Thus the optimization function for this approach is:

$$O^{B-WMRMF} = \sum_{r=1}^{M} \theta_r \sum_{(s,i)\in R_r} ((R_r)_{si} - (b_s + b_i + w_{r1s} w_{r2i}^T))^2 + \lambda(\sum_{j=1}^{N} \| W_j \|_F^2 + b_s^2 + b_i^2) \tag{8}$$

and the student performance prediction is made as:

$$\hat{p} = b_s + b_t + \sum_{k=1}^{K} w_{sk} h_{ik} \tag{9}$$

A detailed algorithm for this personalized multi relational model (Proposed approach 2) has been given in the coming section. Procedure ENHANCED-WMRMF starts with weight and latent factor initialization. The next set of steps is the iterative stochastic gradient descent process. In each iteration, for a random row of a relation, predicted value is calculated using the dot product of latent factors and the biases. Error is calculated as the difference between the predicted value and the known target value. Using this error, the latent factors and the biases are updated. The iterative process stops when the error between two consecutive iterations reaches below a threshold value(stopping criterion). Once the latent factors and the biases are updated using this algorithm, for any new combination of student-task test data, performance prediction can be made easily using equation 9.

Assuming the algorithm takes n iterations to converge, and as specified in the input section of the algorithm, for a model with N entities, M relations with R as the size of each relation, f number of factors in each latent factor matrix, the time complexity of both the base models(WMRMF and MRMF) and the proposed personalized approach is given as $O(n.M.R.f)$.

Algorithm 1. Personalized Multi relational matrix factorization model.

Input
N : $enitites$
M : $relations$
F : $number\ of\ factors$
θ : $weight$
β : $regularization\ term$
λ : $learn\ rate$
Output
b_s : $student\ bias$
b_t : $task\ bias$
w_j : $latent\ factors\ of\ each\ entity\ j$
procedure ENHANCED- WMRMF($N, M, F, \theta, \beta, \lambda$)
  Initialize weight value θ for each relation.
  Initialize latent factors w_j , $(j \in N)$ for each of the N entities.
  while $stopping\ criterion\ met$ **do**
    for $i = 1\ to\ M$ **do**
      for $j = 1\ to\ size(m)$, $(m \in M)$ **do**
        $Pick\ a\ row\ (p, q)\ from\ relation\ m\ in\ random\ with\ target\ value\ p$
        $\hat{p} = b_s + b_t + w_{r1p} w_{r2q}$
        $error_{pq} = p - \hat{p}$
        $w_{r1p} \leftarrow w_{r1p} + \beta * (\theta_m * error_{pq} * w_{r2q} - \lambda w_{r1p})$
        $w_{r2q} \leftarrow w_{r2q} + \beta * (\theta_m * error_{pq} * w_{r1p} - \lambda w_{r2q})$
        $b_s \leftarrow b_s + \beta * (\theta_m * error_{pq} - \lambda b_s)$
        $b_i \leftarrow b_i + \beta * (\theta_m * error_{pq} - \lambda b_i)$
      end for
    end for
  end while
end procedure

---

4 Experimental Analysis

We implemented the personalized multi relational matrix factorization model on a 4GB RAM, 64 bit Operating System, Intel Core i3 machine. The model implementation is done using Java Version 8. Knowledge Discovery Data Challenge Algebra 2005-2006 dataset has been used. This data set is a click-stream log describing interaction between students and intelligent tutoring system [12]. Initially a preprocessing of the dataset, a one time activity is done using MATLAB 2012b (Version 8.0). In the preprocessing stage, data is read and unique id is assigned to each set of student, task and skill. From the dataset, we take three relations into consideration namely Student-performs-task, task-Requires-Skill and Student-haslearnt-Skill. Average performance of a student for a task is calculated from the dataset and assigned as the target value for Student-performs-task. Target value for Task requires Skill relation is either 1 (requires) or 0 (not requires). Opportunity count which represents the number of chances the student got to learn the skill forms the target value of Student-haslearnt-Skill. This value is implicitly learnt from the dataset. Thus in the pre-processing stage, the relation between entities is represented with id value assigned and corresponding target value is decided. Hence this is the most time consuming phase as compared to the actual training and prediction phase in case of multi relational models and can take hours to complete. Next, we start the training procedure - Iterative updating of the factors and bias term in order to minimize the error between the predicted and the target values as discussed in the algorithm. For experimentation, the error measure is taken as RMSE and MAE. Once we obtain the optimized latent factors and bias term, prediction can be made for any data. The global average value has been used for dealing with cold start problems. Since stochastic gradient descent is used for training, we observe a fast convergence of latent factors and bias term. Further, prediction can be done in real time.

5 Results

Dataset was divided into chunks of different sizes and experimentation was performed for the base models and the proposed approach. As explained in the previous section, Root Mean Square Error, RMSE (fig 1), Mean absolute error, MAE (fig 2) and accuracy (fig 3) were calculated for each dataset chunks considered. Graph plots giving comparison of different models are given below. Below three figures gives comparison of base models (`MRMF` and `WMRMF`) and the proposed approaches (`WMRMF_bias_avg` and `WMRMF_bias`). We notice a drastic change in accuracy while the bias term was also considered for multi relational model. Also, accuracy of model increase slightly while neglecting the global average term for performance prediction of existing students and tasks.

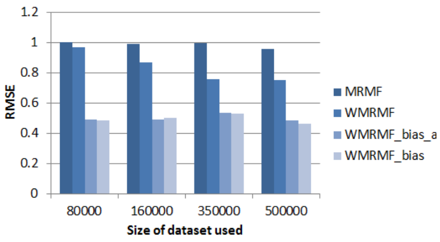

Fig. 1 Student Performance Prediction RMSE

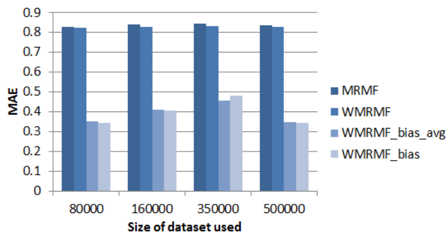

Fig. 2 Student Performance Prediction MAE

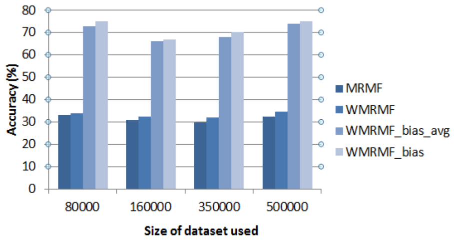

Fig. 3 Student Performance Prediction Accuracy

6 Conclusion and Future Work

Matrix factorization models prove very effective for student performance prediction. In the past multiple relationships amongst the entities have been exploited in order to improvise the state-of-the-art systems leading to Multi-Relational Matrix Factorization (MRMF) and Weighted MRMF. In this paper, we propose to enhance the prediction accuracy by developing a personalized Weighted MRMF model. For this two approaches have been proposed. First approach considers both global average and bias terms along with the factors in predicting student performance. Second approach considers only bias with the latent factor products for performance prediction. Through experimental analysis with different dataset sizes from KDD Cup Challenge, we show that both the models achieve higher prediction accuracy and lower RMSE when compared to the base models, although considering only biases and neglecting global average prove to achieve the best results.

Training data is a snapshot taken at a particular time. As new data for prediction comes in, it could be used for online updating of the system to give out better prediction results [10]. Also, considering any other relation other than the ones considered in this paper may lead to better accuracy.

Acknowledgments This work derives direction and inspiration from the Chancellor of Amrita University, Sri Mata Amritanandamayi Devi.We would like to thank Dr.M Ramachandra Kaimal, Head, Department of Computer Science and Dr. Bhadrachalam Chitturi, Associate Professor, Department of Computer Science, Amrita University for their valuable feedback.

References

1. Krohn-Grimberghe, A., Drumond, L., Freudenthaler, C., Schmidt-Thieme, L.: Multi-relational matrix factorization using bayesian personalized ranking for social network data. In: Proceedings of thefifth ACM international conference on Web search and data mining-WSDM 2012
2. Bekele, R., Menzel, W.: A bayesian approach to predict performance of a student (bapps): a case with ethiopian students. In: Proceedings of the International Conference on Artificial Intelligence and Applications, Vienna, Austria, vol. 27, pp. 189–194 (2005)
3. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction **4**, 253–278 (1995)
4. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review **51**(3), 455–500 (2009)
5. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. In: IEEE Computer Society Press, vol. 42, pp. 30–37 (2009). 38, 40, 41, 45
6. Lippert, C., Weber, S.H., Huang, Y., Tresp, V., Schubert, M., Kriegel, H.P.: Relation prediction in multi-relational domains using matrix factorization. In: NIPS 2008 Workshop: Structured Input - Structured Output (2008). 56, 57, 60, 61, 64
7. Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G., Punch, W.F.: Predicting student performance: an application of data mining methods with an educational web-based system. In: The 33rd IEEE Conference on Frontiers in Education (FIE 2003), pp. 13–18 (2003). 27
8. Nedungadi, P., Remya, M.S.: Predicting students' performance on intelligent tutoring system personalized clustered BKT (PC-BKT) model. In: Frontiers in Education Conference (FIE). IEEE (2014)
9. Nedungadi, P., Smruthy, T.K.: Enhanced higher order orthogonal iteration algorithm for student performance prediction. In: International Conference on Computer and Communication Technologies (2015). AISC Springer
10. Rendle, S., Schmidt-Thieme, L.: Online-updating regularized kernel matrix factorization models for large-scale recommender systems. In: Proceedings of the ACM conference on Recommender Systems (RecSys 2008), pp. 251–258. ACM, New York (2008). 40
11. Singh, A.P., Gordon, G.J.: Relational learning via collective matrix factorization. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), KDD 2008, pp. 650–658. ACM, NewYork (2008). 56, 58, 60
12. Stamper, J., Niculescu-Mizil, A., Ritter, S., Gordon, G.J., Koedinger, K.R.: Algebra 2005–06 Challenge dataset from KDD Cup 2010 Educational Data Mining Challenge (2010)
13. Thai-Nghe, N., Drumond, L., Horvath, T., Nanopoulos, A., Schmidt-Thieme, L.: Matrix and tensor factorization for predicting student performance. In: Proceedings of the 3rd International Conference on Computer Supported Education (CSEDU 2011) (2011a)
14. Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., Schmidt-Thieme, L.: Recommender system for predicting student performance. In: Proceedings of the ACM RecSys 2010 Workshop on Recommender Systems for Technology Enhanced Learning (RecSys-TEL 2010), vol. 1, pp. 2811–2819. Elsevier's Procedia Computer Science (2010c)

15. Thai-Nghe, N., Drumond, L., Horvath, T., Krohn-Grimberghe, A., Nanopoulos, A., Schmidt-Thieme, L.: Factorization techniques for predicting student performance. In: Santos, O.C., Boticario, J.G. (eds.) Educational Recommender Systems and Technologies: Practices and Challenges (ERSAT 2011). IGI Global (2011)
16. Thai-Nghe, N., Drumond, L., Horvath, T., Schmidt-Thieme, L.: Multi-relational factorization models for predicting student performance. In: Proceedings of the KDD 2011 Workshop on Knowledge Discovery in Educational Data (KDDinED 2011) (2011c)