Rank-Based Tensor Factorization for Student Performance Prediction

Thanh-Nam Doan
Department of Computer Science
University at Albany - SUNY
Albany, NY
tdoan@albany.edu

Shaghayegh Sahebi
Department of Computer Science
University at Albany - SUNY
Albany, NY
ssahebi@albany.edu

ABSTRACT

One of the essential problems, in educational data mining, is to predict students' performance on future learning materials, such as problems, assignments, and quizzes. Pioneer algorithms for predicting student performance mostly rely on two sources of information: students' past performance, and learning materials' domain knowledge model. The domain knowledge model, traditionally curated by domain experts, maps learning materials to concepts, topics, or knowledge components that are presented in them. However, creating a domain model by manually labeling the learning material can be a difficult and time-consuming task. In this paper, we propose a tensor factorization model for student performance prediction that does not rely on a predefined domain model. Our proposed algorithm models student knowledge as a soft membership of latent concepts. It also represents the knowledge acquisition process with an added rank-based constraint in the tensor factorization objective function. Our experiments show that the proposed model outperforms state-of-the-art algorithms in predicting student performance in two real-world datasets, and is robust to hyper-parameters.

Keywords

student modeling, predicting student performance, tensor factorization

1. INTRODUCTION

The popularity of online learning services and massive open online courses has led to extensive growth in the amount of student activity and learning data. As the number of students and learning materials increase in these online systems, the need for automatic sense-making from this data, educational data mining, becomes more evident. One of the important tasks in educational data mining is accurately predicting students' performance (PSP). PSP can be used in early detection of high-risk students that may fail or quit a class, in class evaluation and course planning activities, and in learning material recommendation to students.

Thanh-Nam Doan and Shaghayegh Sahebi "Rank-Based Tensor Factorization for Predicting Student Performance" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 288 - 293

Many successful PSP techniques aim to predict students' performance in a problem by modeling their state of knowledge in different concepts required by that problem. To do this, pioneer and recent PSP techniques rely on the availability of a domain knowledge model that maps problems to concepts [19, 5, 25]. However, given the vast scope of learning materials in today's online learning systems, such domain knowledge models may not be available. Ideally, a PSP model should be able to work without requiring such a predefined map.

Additionally, a successful data mining model for PSP should be capable of considering specific characteristics of student learning process: (a) that students gain their knowledge on concepts over time, by practicing different problems, (b) that they may forget some of the gained knowledge, (c) that this knowledge gain is a gradual process, and (d) that learning can happen differently for different students in different problems and different times. Finally, to provide better insight to students and teachers, such a model should also be interpretable considering these characteristics. Previous research in the literature only cover some of the limitations above.

In this paper, we propose a student performance prediction model, Ranked-Based Tensor Factorization (RBTF), considering all the above requirements. To model student sequences on problems, we represent their scores over time as a three-dimensional tensor. To avoid the need for a predefined domain knowledge model, we propose a tensor factorization model for PSP, that maps problems and student knowledge in a lower-dimensional "latent" concept space. Representing student knowledge in this lower-dimensional space leads to a soft-membership approach that provides more flexibility by avoiding strict assignment of student knowledge to discrete "knowledge states". By learning student, problem, and time-based biases in this model we take into account the differences between students, problems, and times in the learning process. To capture the gradual learning requirement, we impose a rank-based constraint on student knowledge variables, that allows for occasional forgetting of concepts, but imposes a generally positive learning trend.

In our experiments, we study the proposed model in comparison with two state-of-the-art baseline PSP algorithms, on two real-world datasets. Our experiments show that our model performs better than both baselines in the task of predicting student performance. We experiment with the performance and sensitivity of our model with various hyper-parameters.

Paper Outline. The remaining of the paper is organized as follows. Section 2 provides a brief literature review of the related work.Section 3 describes our model (RBTF) and the parameter learning steps. Section 4 evaluates extensively RBTF and other baselines on two real datasets. Lastly, Section 5 concludes the paper and suggests some directions for future works.

2. RELATED WORK

Many pioneer solutions to the problem of predicting student performance are based on either regression models [19] or Bayesian knowledge tracing (BKT) [5]. Regression-based models, such as performance factor analysis (PFA), try to predict students' performance using a pre-defined domain model that maps learning material to knowledge components [19]. PFA, which is based on learning factor analysis [4], takes into account prior successes and failures of a student on knowledge components associated with the current problem.

BKT is a constrained two-state hidden Markov model that models student knowledge in each knowledge component (KC) as two binary states: "known" and "unknown". It learns the probability of transitioning between these two states, and probabilities of students' success and failure in each KC, given their state of knowledge. Despite being successful in PSP for certain datasets, this model, in its original form, does not consider continuous states of knowledge or soft membership to knowledge states. Moreover, BKT does not capture the relationships between KCs, and is not personalized for individual student. Additionally, BKT also relies on a pre-defined domain model. Recently, new BKT-based models aim to address some of these problems [2, 9, 29]. For example, Pardos and Heffernan has addressed BKT's non-personalized modeling in [18, 17]. Song et al. proposed PSFK in [25] to address PSP when students encounter a knowledge component for the first time. But, these models rely on labeled problem knowledge components or concepts. Later, Gonzalez-Brenes and Mostow proposed a topical hidden Markov model that jointly learns the domain model and predicts student performance [10, 8]. However, this model has two restricting assumptions: that at each attempt, the student works on one skill of a problem, and that the students do not forget any acquired skills.

Recently, other approaches inspired by recommender systems' research and factorization models have been used for PSP. Despite being successful, these approaches are not tailored for the educational data mining problems specifically since they do not explicitly model student learning as a learning gain process. The matrix-factorization approaches in this area do not consider student sequences and only rely on a snapshot of student performance. For example, Thai-Nghe and Schmidt-Thieme proposed a multi-relational factorization student model that considers multiple relations between students and tasks, but does not consider student sequences [27]. Later, Nedungadi and Smruthy proposed a similar multi-relational matrix factorization approach exploring the effect of modeling biases [16]. Sahebi et al. also proposed another multi-relational learning approach that learned student performance according to canonical correlation analysis [22]. Non-negative matrix factorization has been used to improve performance predictions [28]. Pero et al. compared collaborative filtering techniques for the task of PSP in a small dataset [20]. Elbadrawy et al. predict student performance using their interactions with the learning management system to achieve a higher accuracy [7].

Some other recommender system-based approaches consider student sequence, but do not explicitly model knowledge gain in students. For example, Thai-Nghe et al. explored different factorization models, including tensor factorization, to predict student performance [26]. Sahebi et al. [23] studied educational data mining methods, such as PFA and BKT, with matrix and tensor factorization approaches, from the recommender systems literature, for PSP. Almutairi et al. have used tensor and coupled-matrix factorization to predict course-based student performance [1]. However, their tensor decomposition models do not explicitly model students' knowledge gain.

Although there have been some promising research on PSP that consider student sequence without requiring a domain model, these approaches have been limited. For example, SPARse Factor Analysis (SPARFA) by Lan et al. that uses Kalman filters to jointly learn the domain model, student knowledge, and the underlying question difficulties, can be very expensive to learn due to having a big state space [12]. Sahebi et al. have proposed a feedback-driven tensor factorization algorithm that can model student gradual knowledge acquisition [24]. But, their model has a strict constraint that does not allow for forgetting the concepts by students. Lindsey et al. proposed a non-parametric Bayesian technique that can refine the expert-labeled skills. However, they simplify the problem by finding coarse-grained skills as they restrict each problem to have exactly one skill [14]. In this paper, we propose a tensor factorization model for predicting student performance that does not require domain knowledge, models problems and student knowledge as soft-membership of latent concepts, and can model student sequences and gradual knowledge increase.

3. RANK-BASED TENSOR FACTORIZATION (RBTF)

Here we present our model, rank-based tensor factorization, by which we aim to predict students' performance in problems, considering their performance sequence and knowledge growth. Our proposed model is inspired by the recommender systems domain. Our choice of a recommender systems-based model was because of two main reasons: a) student performance similarities, and b) problem similarities. First, we consider that students with similar knowledge levels will perform similarly in solving problems. Second, we assume that a student will have similar performance on two problems with similar concepts. Recommender-based factorization models consider these two expectations. However, as discussed in the introduction section, a successful student model needs to include additional considerations. One of which is that knowledge gain is a gradual process for students, which happens over time. As students interact with learning materials, such as problems, they learn from them.

To represent this time-based process, we model students' activity sequences as a series of attempts on problems. For the student performance data to be represented according to these assumptions, we represent student sequences in a three-dimensional tensor ($\mathcal{Y}$), that has the student, problem, and time (attempt) dimensions. Each cell $y_{a,s,p}$ in this tensor represents student s's score in problem p, that she has chosen to study at attempt a.

The core idea of the aforementioned assumptions is the notion of "concepts": gradual learning can be viewed as gaining knowledge in course concepts; student knowledge-based similarities are based on how much they mastered each of the concepts; and problem similarities are defined on how their represented concepts are shared. However, in many online educational systems, concepts are undefined and difficult to measure. In these systems, there are no "observed" features defined as problem concepts or knowledge components. Hence, we propose to discover shared "latent" features between students and problems as representatives for the notion of concepts. We model each problem as a vector of k latent concepts, that shows the importance of each latent concept in that problem. Also, we model each student's knowledge at any time point as another vector of the same latent concepts.

We assume that a student s's performance on problem p at time a is a result of her existent knowledge in the latent concepts required by the problem. Accordingly, we model estimated student score $\hat{y}_{a,s,p}$ as a dot product between problem's latent concept vector $\boldsymbol{q}_p$ and student's knowledge in those concepts $\boldsymbol{t}_{a,s}$:

$$\hat{y}_{a,s,p} \approx \boldsymbol{t}_{a,s}.\boldsymbol{q}_p \tag{1}$$

To maintain the interpretability of our model, we enforce latent variables in $\boldsymbol{q}_p$ to be non-negative. Here, by choosing the number of concepts (k) less than the number of problems and students, we are representing students and problems in a lower-dimensional latent space that can better capture student and problem similarities (our second and third assumptions). However, the model in Equation 1 does not consider differences in factors such as student ability, problem difficulty, or student cohort strength. For example, students' average score in a more difficult problem is expected to be less than their average score in an easier problem. To address this issue, we add student, problem, and attempt biases (b_s, b_p, and b_a), in addition to an overall cohort bias (μ) to our model:

$$\hat{y}_{a,s,p} \approx \boldsymbol{t}_{a,s}.\boldsymbol{q}_p + b_s + b_p + b_a + \mu \tag{2}$$

To learn the parameters of this problem ($\mathcal{T}$, Q, b_s, b_p, b_a, and μ) we minimize the objective function in Equation 3. The first component calculates the squared difference between observed student scores and estimated student scores. The last three components are for regularizing biases, student knowledge, and problem concepts for generalizability purposes.

$$\begin{aligned}\mathscr{L}_1 = &\sum_{a,s,p} (\hat{y}_{a,s,p} - y_{a,s,p})^2 \\ &+ \lambda(b_s^2 + b_p^2 + b_a^2) + \lambda_1 \left\|\boldsymbol{t}_{a,s}\right\|^2 + \lambda_2 \left\|\boldsymbol{q}_p\right\|^2\end{aligned} \tag{3}$$

The model in Equation 2 does not address our gradual learning assumptions for students. To capture this gradual learning, we can assume that a student's knowledge ($\boldsymbol{t}_{a,s}$) increases over time. But, we should also note that this knowledge increase depends on the problems that the student selects to solve and the concepts presented in them. As a result, we can translate this knowledge increase as an increase in estimated student scores in problems ($\boldsymbol{t}_{a,s}.\boldsymbol{q}_p$). In other words, we expect that student s's predicted scores at attempt a to be larger than her scores at attempt $a-1$:

$$\boldsymbol{t}_{a,s}.\boldsymbol{q}_p - \boldsymbol{t}_{a-1,s}.\boldsymbol{q}_p \geq 0$$

In reality, this knowledge increase can be non-monotonic. For example, a student may forget some concepts after a while. For this reason, we propose to use a rank-based model for student knowledge gain, that allows knowledge loss to happen for students, but penalizes it. Using this rank-based model, we aim to maximize the difference between the aggregation of all students' scores on all questions at each attempt versus the attempts before that. Hence, we would like to maximize $\mathscr{L}_2$ in Equation 4. Here, $\sigma(\cdot)$ is the sigmoid function, defined as $\sigma(x) = 1/(1+e^{-x})$. Sigmoid function is selected because of its superiority in rank-based recommendation systems [21, 6]. The term $\log(\sigma(\boldsymbol{t}_{a,s}\boldsymbol{q}_p - \boldsymbol{t}_{j,s}\boldsymbol{q}_p))$ means that for attempt a of student s, the *ranking* of s's score at a is higher than the one of s at j with $j < a$.

$$\mathscr{L}_2 = \sum_{j=1}^{a} \sum_{s} \sum_{p} \log(\sigma(\boldsymbol{t}_{a,s}\boldsymbol{q}_p - \boldsymbol{t}_{j,s}\boldsymbol{q}_p)) \tag{4}$$

To capture the dynamics between all assumptions, we combine the minimization of $\mathscr{L}_1$ in Equation 3 and maximization of $\mathscr{L}_2$ in Equation 4. Our final objective is to minimize the loss function in Equation 5. The hyper-parameter ω is to control the relative strictness of knowledge increase versus the importance of having a more accurate estimate of student performance.

$$\begin{aligned}\mathscr{L} = &\sum_{a,s,p} (\hat{y}_{a,s,p} - y_{a,s,p})^2 + \lambda_1 \left\|\boldsymbol{t}_{a,s}\right\|^2 + \lambda_2 \left\|\boldsymbol{q}_p\right\|^2 \\ &+ \lambda(b_s^2 + b_p^2 + b_a^2) - \omega \sum_{j=1}^{a} \sum_{s} \sum_{p} \log(\sigma(\boldsymbol{t}_{a,s}\boldsymbol{q}_p - \boldsymbol{t}_{j,s}\boldsymbol{q}_p))\end{aligned} \tag{5}$$

Learning the Parameters: By using stochastic gradient descent algorithm to minimize $\mathscr{L}$, we find student knowledge of each latent concept at any point, the importance of each latent concept in each problem, and estimation of student score in each problem at any attempt. Recall that the parameters the we want to infer are $\mathcal{T}$, Q, b_s, b_p, b_a, and μ. For the cohort bias μ, we assign the average score of all students on all problems [11], i.e. $\mu = \frac{\sum_{a,s,p} y_{a,s,p}}{\sum_{a,s,p} \mathcal{I}(a,s,p)}$ where $\mathcal{I}(a,s,p)$ is an indicator function returning 1 if the tuple (a,s,p) is in our training set; otherwise 0.

4. EXPERIMENTS

In the following, we evaluate our model in comparison with two state-of-the-art methods in the task of predicting student performance. Further, we analyze how our solution models students' learning process by looking at students' knowledge gain in course concepts. Eventually, we experiment on RBTF's sensitivity to various hyper-parameter settings.

4.1 Dataset and Experiment Setup

For experiments, we use the Canvas network dataset[1] which is available online [3]. Canvas Network hosts many freely available open online courses. In addition to learning modules, each course can have different types of assignments, discussions, and quizzes. In this platform, participants are not limited to a specific sequence of learning material or assignments. The dataset is anonymized such that student IDs, course names, discussion contents, submission contents, or course contents are not available.

Dataset	#students	#problems	#attempts	Avg. attempts
Course 1	531	91	87	29.92
Course 2	2597	32	30	12.73

Table 1: Dataset Statistic.

We select two courses in Canvas and denote them as Course 1 and Course 2. The selected courses have the most number of quizzes in the whole dataset. We consider each quiz as a problem in our model. Quizzes are graded between zero and a maximum possible score. For consistency, we normalize the quiz grades between zero and one. Table 1 shows the statistics of these two courses. As shown in the table, Course 2 has more students but less number of problems and attempts than Course 1.

The data of each course is represented as a list of tuples *(attempt, student id, quiz id, grade)*. We randomly split 80% of tuples for training and the remaining (i.e. 20%) for testing.

Hyper-parameter Setting: In the performance prediction experiments (Section 4.2), we set $\omega = 0.5$, $\lambda_1 = \lambda_2 = 0.1$ and regularization of bias $\lambda = 0.001$. The number of concepts is set to 3.

4.2 Student Performance Prediction

In this section, we compare the prediction performance of RBTF with other baselines to evaluate the prediction ability of RBTF.

Baselines: To compare the prediction performance, we employ the following two baselines:

- Feedback-Driven Tensor Factorization (FDTF) [24]: It is a tensor factorization model specifically tailored to predict students' performance. It considers students' gradual learning process. However, the assumption of hard constraint on knowledge increase in students limits its modeling capacity. Also, it does not include biases and does not allow for the concepts to be forgotten by students.

- SPARse Factor Analysis (SPARFA) [13]: SPARFA is a probabilistic factor analysis approach that calculates the probability of a student's correct response to a problem. It does not require a predefined domain knowledge model. However, it does not consider students' sequences. To adapt it to our problem, we use the probability scores instead of the predicted student grade.

Metrics: We use two measures to evaluate the performance prediction task. Since our main goal is to predict student scores or grades, we would like to measure how close our predictions are to the actual student scores. To do this, we use the root mean squared error (RMSE). The lower the value of RMSE, the better the model.

Since many performance prediction models focus on predicting students' success and failure as a binary value, instead of their score [13, 5], we also employ accuracy for performance comparison. To do this, we regard scores greater than 0.5 as success and the rest as failure. Unlike RMSE, the higher the value of accuracy, the better the model.

Dataset	RMSE			Accuracy		
	RBTF	FDTF	SPARFA	RBTF	FDTF	SPARFA
Course 1	**0.12**	0.27	0.59	**92.5%**	85.2%	81.7%
Course 2	**0.2056**	0.2116	0.567	**95.24%**	92.8%	87.41%

Table 2: Prediction Performance.

Results: Table 2 shows the prediction performance of our model (RBTF) and the two baselines (FDTF and SPARFA) on the two datasets. As we can see, both tensor factorization models (RBTF and FDTF) perform better than SPARFA in both courses. This shows the importance of considering student sequences in predicting their performance. Also, we can see that RBTF performs better than FDTF in both courses. This shows that, even though modeling sequential knowledge increase in students is important, this increase should not be strictly monotonic and should be flexible to allow for occasional forgetting of concepts.

4.3 Hyper-parameter Sensitivity Analysis

In this section, we study RBTF's sensitivity to hyper parameter values. First, we experience on the balance between training error on student performance fitting ($\mathscr{L}_1$ in Equation 3) versus modeling student knowledge increase ($\mathscr{L}_2$ in Equation 4) on the generalizability of our model. To do this, we measure the test error by varying hyper-parameter ω, that controls this balance in Equation 5. Then, we capture the effect of the number of concepts on RBTF's performance by varying k in Equation 5 and measuring its error on test data.

[1]`http://canvas.net`

	ω				
Dataset	0.01	0.25	0.5	0.75	1.0
Course 1	0.191	0.128	**0.12**	0.137	0.141
Course 2	0.233	0.2064	**0.2056**	0.2154	0.2224

Table 3: RMSE with different value of ω and number of concept is 3.

Sensitivity to ω: Recall that ω controls the trade-off between having an accurate estimation of student performance and the constraint of knowledge increase. A larger value of ω, imposes more contribution of knowledge increase constraint to the performance of RBTF, and a smaller value of ω dictates a stricter fit of student performance to the training data. We use different values of ω from 0 to 1 and measure RBTF's RMSE corresponding to these values. For other parameters, we use the default values mentioned in Section 4.1. Table 3 presents the performance of RBTF with different values of ω on the two datasets. From the table, we observe that $\omega = 0.5$ yields the best performance of RBTF and it is consistent for the two datasets. However, the results from Course 2 dataset is more sensitive to the changes in ω. One reason for this can be the smaller number of attempts and more sparsity of Course 2 dataset, compared to Course 1 dataset, that can lead to easier overfitting to training data.

	k			
Dataset	3	5	10	15
Course 1	**0.12**	0.122	0.127	0.128
Course 2	**0.2056**	0.206	0.2065	0.2065

Table 4: RMSE with different value of number of concepts and $\omega = 0.5$.

Sensitivity to k: Recall that, in our model, concepts are latent lower-dimensional representations of student performance and problems over attempts. They can be used to model the similarity between students and problems. To measure the effect of the number of concepts k, we tune the value of k while using the default values for other parameters (see Section 4.1). We measure the RMSE of RBTF by changing k. Table 4 shows the results. From the table, we observe that increasing the value of k makes RBTF perform slightly worse. This finding is consistent in both datasets. However, RBTF is relatively robust to k as this increase in error is minor.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel rank-based tensor factorization method (RBTF), which is able to predict the performance score of students by considering the gradual learning of students as a ranking problem. Our model has the flexibility to present student knowledge as a soft-membership of latent concepts, only requires activity sequences of students, and discovers individualized student knowledge model including biases. Our extensive evaluations show that RBTF outperforms state-of-the-art baselines in both root mean square error and accuracy measures. Also, we show our models robustness to hyper-parameters by experimenting the balance between knowledge ranking and performance fitting parts of the model, and by varying the number of latent concepts.

There are several directions to extend this research work further. In this work, we experiment on performance prediction within the same course. This model can be used to experiment on between-course performance predictions. Another application of our model is to detect knowledge gaps in students and recommend useful learning materials to them. Moreover, contingent on the availability of a domain knowledge model, this work can be extended to improve the existing domain knowledge model. Recent studies show that order and length of students' activities are essential for understanding students' performance [15]. So, integrating these features can enhance the prediction performance of our model.

Acknowledgement. This material is based upon work partially supported by the National Science Foundation under Grant No. 1755910.

6. REFERENCES

[1] F. M. Almutairi, N. D. Sidiropoulos, and G. Karypis. Context-aware recommendation-based learning analytics using tensor and coupled matrix factorization. *IEEE Journal of Selected Topics in Signal Processing*, 11(5):729–741, 2017.

[2] R. S. Baker, A. T. Corbett, and V. Aleven. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *International conference on intelligent tutoring systems*, pages 406–415. Springer, 2008.

[3] Canvas-Network. Canvas network courses, activities, and users (4/2014 - 9/2015) restricted dataset, 2016.

[4] H. Cen, K. Koedinger, and B. Junker. Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Intelligent Tutoring Systems*, pages 164–175. Springer, 2006.

[5] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

[6] T.-N. Doan and E.-P. Lim. Modeling location-based social network data with area attraction and neighborhood competition. *Data Mining and Knowledge Discovery*, 33(1):58–95, 2019.

[7] A. Elbadrawy, R. S. Studham, and G. Karypis. Collaborative multi-regression models for predicting students' performance in course activities. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 103–107. ACM, 2015.

[8] J. González-Brenes. Modeling skill acquisition over time with sequence and topic modeling. In *Artificial Intelligence and Statistics*, pages 296–305, 2015.

[9] J. González-Brenes, Y. Huang, and P. Brusilovsky. General features in knowledge tracing to model multiple subskills, temporal item response theory, and expert knowledge. In *The 7th International Conference on Educational Data Mining*, pages 84–91. University of Pittsburgh, 2014.

[10] J. Gonzalez-Brenes and J. Mostow. What and when do students learn? fully data-driven joint estimation of cognitive and student models. In *Educational Data Mining 2013*, 2013.

[11] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, Aug. 2009.

[12] A. S. Lan, C. Studer, and R. G. Baraniuk. Time-varying learning and content analytics via sparse factor analysis. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 452–461. ACM, 2014.

[13] A. S. Lan, A. E. Waters, C. Studer, and R. G. Baraniuk. Sparse factor analysis for learning and content analytics. *The Journal of Machine Learning Research*, 15(1):1959–2008, 2014.

[14] R. V. Lindsey, M. Khajah, and M. C. Mozer. Automatic discovery of cognitive skills to improve the prediction of student learning. In *Advances in neural information processing systems*, pages 1386–1394, 2014.

[15] M. Mirzaei, S. Sahebi, and P. Brusilovsky. Annotated examples and parameterized exercises: Analyzing students' behavior patterns. In *International conference on artificial intelligence in education*, 2019.

[16] P. Nedungadi and T. Smruthy. Personalized multi-relational matrix factorization model for predicting student performance. In *Intelligent Systems Technologies and Applications*, pages 163–172. Springer, 2016.

[17] Z. A. Pardos and N. T. Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 255–266. Springer, 2010.

[18] Z. A. Pardos and N. T. Heffernan. Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. *Journal of Machine Learning Research W & CP*, 2010.

[19] P. I. Pavlik Jr, H. Cen, and K. R. Koedinger. Performance factors analysis–a new alternative to knowledge tracing. *Online Submission*, 2009.

[20] Š. Pero and T. Horváth. Comparison of collaborative-filtering techniques for small-scale student performance prediction task. In *Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering*, pages 111–116. Springer, 2015.

[21] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.

[22] S. Sahebi and P. Brusilovsky. Student performance prediction by discovering inter-activity relations. *Educational Data Mining*, pages 87–96, 2018.

[23] S. Sahebi, Y. Huang, and P. Brusilovsky. Predicting student performance in solving parameterized exercises. In *Intelligent Tutoring Systems*, pages 496–503. Springer, 2014.

[24] S. Sahebi, Y. Lin, and P. Brusilovsky. Tensor factorization for student modeling and performance prediction in unstructured domain. In *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*, pages 502–506, 2016.

[25] Y. Song, Y. Jin, X. Zheng, H. Han, Y. Zhong, and X. Zhao. Psfk: a student performance prediction scheme for first-encounter knowledge in its. In *International Conference on Knowledge Science, Engineering and Management*, pages 639–650. Springer, 2015.

[26] N. Thai-Nghe, T. Horváth, and L. Schmidt-Thieme. Factorization models for forecasting student performance. In *Educational Data Mining 2011*, 2010.

[27] N. Thai-Nghe and L. Schmidt-Thieme. Multi-relational factorization models for student modeling in intelligent tutoring systems. In *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*, pages 61–66. IEEE, 2015.

[28] K. Xu, R. Liu, Y. Sun, K. Zou, Y. Huang, and X. Zhang. Improve the prediction of student performance with hint's assistance based on an efficient non-negative factorization. *IEICE TRANSACTIONS on Information and Systems*, 100(4):768–775, 2017.

[29] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon. Individualized bayesian knowledge tracing models. In *International conference on artificial intelligence in education*, pages 171–180. Springer, 2013.