The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Contextual Collaborative Filtering for Student Response Prediction in Mixed-Format Tests

Shumin Jing
University of Iowa
Iowa City, IA 52246
shumin-jing@uiowa.edu

Sheng Li
Adobe Research
San Jose, CA 95110
sheli@adobe.com

Abstract

The purpose of this study is to design a machine learning approach to predict the student response in mixed-format tests. Particularly, a novel contextual collaborative filtering model is proposed to extract latent factors for students and test items, by exploiting the item information. Empirical results from a simulation study validate the effectiveness of the proposed method.

As large-scale assessment data become available in recent years, especially in the online education scenario (e.g., massive open online course, MOOC), a scalable approach for assessing student ability is highly demanded. Traditional approaches like Item Response Theory (IRT) have been widely applied for estimating student ability (Embretson and Reise 2013), but they still have some limitations. For instance, the three parameter logistic (3PL) model only utilizes three parameters to characterize each test item, which might be insufficient in large-scale settings. Machine learning, as an interdisciplinary field of computer science and statistics, has achieved impressive performance in many domains. Given sufficient training data, machine learning approaches optimize the model parameters, and generate predictive results for specific tasks. As an effective machine learning approach, collaborative filtering can be used to predict the item response for dichotomous test items (Bergner et al. 2012). However, to the best of our knowledge, the response prediction for mixed-format tests using machine learning approach has not been investigated before. The primary goal of this study is to design a machine learning approach for item response prediction in mixed-format tests, and evaluate its performance on simulated data. Particularly, the contributions of this work include:

- We design a novel contextual collaborative filtering (CCF) approach to predict item responses in mixed-format tests. CCF can take advantages of the item characteristics estimated by IRT models.
- We evaluate the model performance with different test length, sample size, and proportions for item types in mixed-format tests.

Methodology

Item response theory is powerful in characterizing the test items. For instance, the classical 3PL model uses 3 parameters to quantize the difficulty, discrimination, and guessing. On the other hand, collaborative filtering, especially the matrix factorization based methods, has shown impressive performance in response prediction (Li, Kawale, and Fu 2015). Motivated by these two techniques, we propose a novel matrix factorization method named contextual collaborative filtering (CCF), by leveraging the characteristics of test items to boost the response prediction performance.

Let $X \in \mathbb{R}^{N \times T}$ denote an item response matrix, from N students to T items. Each row in X corresponds to a student, and each column corresponds to an item. The element X_{ij} denotes the response from the i-th student to the j-th test item. X could be very sparse, as we might not know all the responses from students to test items. In mixed-format tests, $X_{ij} \in [0,\ C]$, where C is the maximum number of categories in the response. Matrix factorization is an effective collaborative filtering approach (Koren and Bell 2015), which decomposes the response matrix X into two latent factors U and V, i.e., $X \approx UV$, $U \in \mathbb{R}^{N \times K}$, $V \in R^{K \times T}$, where K is the dimensionality of latent factors and it's usually much smaller than N and T. The objective function of standard matrix factorization is:

$$\begin{aligned} \arg\min_{U,V} f(U,V) = \|I \odot (X - UV^{\mathrm{T}})\|_{\mathrm{F}}^2 \\ + \lambda(\|U\|_{\mathrm{F}}^2 + \|V\|_{\mathrm{F}}^2), \end{aligned} \tag{1}$$

where I is an indictor matrix with $I_{ij} = 1$ if X_{ij} is a valid value and 0 otherwise, $\odot$ denotes the Hadamard product (i.e., element-wise product), and λ is a trade-off parameter. The first term in (1) denotes the approximation errors, and the last two terms are regularizations used to prevent overfitting.

Each row in U is a compact vector, which can be considered as parameters (or features) for characterizing the corresponding student. Similarly, each column in V can be regarded as the parameters for the corresponding item. Thus, this approach offers more flexibility than IRT, in terms of modeling the response patterns of students. However, a major limitation of model (1) is that it doesn't exploit additional information from the test items. For instance, the test items could be characterized by a few *item parameters*, which enables us to model the similarity among all test items. As each column in V denotes latent representations of test items, it is reasonable to assume that similar test items should have similar latent representations. First, we construct a graph G by using the item parameters which are estimated by an IRT model. It's standard procedure in educational measurement. Second, we design a new constraint to enforce that similar items should have similar representations. Formally, we have the objective function of contextual collaborative filtering (CCF) as:

$$\begin{aligned}\arg\min_{U,V} f(U,V) = \|I \odot (X - UV^{\mathrm{T}})\|_{\mathrm{F}}^2 + \alpha f(G,V) \\ + \lambda(\|U\|_{\mathrm{F}}^2 + \|V\|_{\mathrm{F}}^2),\end{aligned} \tag{2}$$

where $f(G,V) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} G_{ij}\|v_i - v_j\|_2^2 = \mathrm{tr}(VLV^{\top})$, L is a Laplacian matrix, $L = \tilde{D} - G$, $\tilde{D}_{ii} = \sum_{j=1}^{n} G_{ij}$, and α is a trade-off parameter.

Eq. (2) shows that CCF is able to take advantages of the test item information, and therefore extract more effective latent factors for items. Standard optimization algorithms, such as stochastic gradient descend (SGD), can be used to learn U and V in (2). Based on the learned latent factors U and V, the predicted response from a student to an item can be estimated by taking the inner product of two corresponding vectors.

Experiments

Settings. As the rationale of IRT models have been widely recognized, both the dichotomous and polytomous IRT models were adopted to generate response data of mixed-format tests. In this study, the R package, *mirt* (Chalmers and others 2012), was used for response data simulation under the two-parameter model (2PL) model and the generalized partial credit model (GPCM). Three test lengths were considered in the study, including 30, 60, and 90. The number of categories was varied from 2 to 5. The proportion of dichotomous items was set to $\frac{1}{3}$. To simulate the large-scale setting, three sample sizes (i.e., the number of students) were considered, including 1000, 2000, and 5000. To evaluate the prediction performance of the proposed method, 80% of responses were randomly selected for training latent factors, while the others were used for testing. The dimensionality of latent factors was fixed to 10. Parameters λ and α were set using the cross-validation strategy on the training set. Following the convention in the collaborative filtering literature, RMSE is chosen as the evaluation criteria in our experiments.

Baselines. The proposed CCF approach is compared with two representative collaborative filtering methods, including probabilistic matrix factorization (PMF) (Mnih and Salakhutdinov 2008) and bayesian probabilistic matrix factorization (BPMF) (Salakhutdinov and Mnih 2008). The dimensionality of latent factors was also fixed to 10, in order to make a fair comparison.

Results. Table 1 shows the RMSE of the proposed method in different settings. We observe that: (1) the RMSE decreases when the sample size is increased from 1000 to 5000; (2) the RMSE decreases when the test length increases; (3) the RMSE decreases when the proportion of dichotomous items increases. It is clear that the parameter estimation becomes more stable when the data size increases, which explains the lower RMSE from longer tests or larger sample size. Moreover, increasing the proportion of dichotomous items would reduce the problem complexity, as binary values are easy to be modeled and predicted.

Table 1: RMSE of response prediction

Test Length	Method	N=1000	N=2000	N=5000
30	PMF	0.3010	0.2917	0.2858
	BPMF	0.3002	0.2910	0.2845
	CCF	0.2914	0.2853	0.2803
60	PMF	0.2847	0.2740	0.2714
	BPMF	0.2805	0.2733	0.2701
	CCF	0.2734	0.2704	0.2685
90	PMF	0.2836	0.2665	0.2637
	BPMF	0.2795	0.2613	0.2615
	CCF	0.2692	0.2602	0.2584

Conclusions

This paper presents a novel collaborative filtering approach for predicting student response in mixed-format tests. The item response matrix can be decomposed to student latent factor and item latent factor. Moreover, the item information is incorporated in the learning process. The learned latent factors can be used to predict missing values in the response patterns. Experimental results show that the proposed method consistently outperforms the baselines in different settings. In addition to response prediction, the proposed method can be applied to other educational measurement tasks, such as ability estimation, classification, etc.

References

Bergner, Y.; Droschler, S.; Kortemeyer, G.; Rayyan, S.; Seaton, D.; and Pritchard, D. E. 2012. Model-based collaborative filtering analysis of student response data: Machine-learning item response theory. *International Educational Data Mining Society*.

Chalmers, R. P., et al. 2012. mirt: A multidimensional item response theory package for the r environment. *Journal of Statistical Software* 48(6):1–29.

Embretson, S. E., and Reise, S. P. 2013. *Item response theory*. Psychology Press.

Koren, Y., and Bell, R. 2015. Advances in collaborative filtering. In *Recommender Systems Handbook*. Springer. 77–118.

Li, S.; Kawale, J.; and Fu, Y. 2015. Deep collaborative filtering via marginalized denoising auto-encoder. In *Proceedings of ACM International on Conference on Information and Knowledge Management (CIKM)*, 811–820. ACM.

Mnih, A., and Salakhutdinov, R. R. 2008. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, 1257–1264.

Salakhutdinov, R., and Mnih, A. 2008. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th International Conference on Machine Learning*, 880–887. ACM.