Quantized Matrix Completion for Personalized Learning

Andrew S. Lan

Rice University

mr.lan@sparfa.com

Christoph Studer

Cornell University

studer@sparfa.com

Richard G. Baraniuk

Rice University

richb@sparfa.com

ABSTRACT

The recently proposed SPARse Factor Analysis (SPARFA) framework for personalized learning performs *factor analysis* on ordinal or binary-valued (e.g., correct/incorrect) graded learner responses to questions. The underlying factors are termed "concepts" (or knowledge components) and are used for *learning analytics (LA)*, the estimation of learner concept-knowledge profiles, and for *content analytics (CA)*, the estimation of question–concept associations and question difficulties. While SPARFA is a powerful tool for LA and CA, it requires a number of algorithm parameters (including the number of concepts), which are difficult to determine in practice. In this paper, we propose *SPARFA-Lite*, a convex optimization-based method for LA that builds on matrix completion, which only requires a *single* algorithm parameter and enables us to automatically identify the required number of concepts. Using a variety of educational datasets, we demonstrate that SPARFA-Lite (i) achieves comparable performance in predicting unobserved learner responses to existing methods, including item response theory (IRT) and SPARFA, and (ii) is computationally more efficient.

Keywords

Personalized learning, learning analytics, content analytics, factor analysis, matrix completion, convex optimization.

1. INTRODUCTION

Recent advances in machine learning propel the design of personalized learning systems (PLSs) that mine learner data (e.g., graded responses to tests or homework assignments) to automatically provide timely feedback to individual learners. Such automated systems have the potential to revolutionize education by delivering a high-quality, personalized learning experience at large scale.

1.1 SPARse Factor Analysis (SPARFA)

The recently proposed SPARse Factor Analysis (SPARFA) framework introduces models and machine learning algorithms for learning and content analytics [16, 17]. *Learning analytics* (LA) stands for the analysis of the knowledge of each learner, while *content analytics* (CA) stands for the analysis of all learning resources, i.e., textbooks, lecture videos, questions, etc. SPARFA analyzes binary-valued (1 for a correct answer and 0 for an incorrect one) or quantized (ordinal-valued, e.g., partial credits) graded responses of N learners to Q questions, in the domain of a course/exam. The key assumption of SPARFA is that the learners' responses to questions are governed by a small number of K ($K \ll N, Q$) latent factors, called "concepts," which are also known as knowledge components [11]. SPARFA performs the joint estimation of (i) question–concept associations, (ii) learner concept knowledge profiles, and (iii) question difficulties, solely from binary-valued graded learner responses. Provided this analysis, SPARFA enables a PLS to provide automated feedback to learners on their individual concept knowledge and to course instructors on the content organization of the analyzed course.

SPARFA, as well as other factor analysis methods, inevitably suffer from the lack of a principled and computationally efficient way to select the appropriate values of the algorithms' parameters, especially the number of latent concepts K. The choice of the number of concepts K is important for two reasons: First, it affects the performance in predicting unobserved learner responses. Second, it determines the interpretability of the estimated concepts, which is key for a PLS to provide human-interpretable feedback to learners. Rule-based intelligent tutoring systems [24] rely on domain experts to manually pre-define the value of K. Such an approach turns out to be labor-intensive and is error prone, which prevents its use for applications in massive open online courses (MOOCs) [20]. SPARFA utilizes cross-validation to select K, as well as all other algorithm parameters [17]. Such an approach is computationally extensive as it requires multiple SPARFA runs to identify appropriate values for all algorithm parameters.

1.2 Contributions

In this work, we propose SPARFA-Lite, a convex optimization-based LA algorithm that automatically selects the number of latent concepts K by analyzing graded learner responses in the domain of a single course/assessment. SPARFA-Lite leverages recent results in quantized matrix completion [15] to analyze quantized graded learner responses, which accounts for the fact that responses are often graded on an ordinal scale (partial credit). Since SPARFA-Lite only has a single algorithm parameter, it has low computational complexity as compared to existing methods such as IRT or conventional SPARFA. We demonstrate the effectiveness of SPARFA-Lite in (i) predicting unobserved learner responses and (ii) performing LA on a variety of real-world educational datasets.

1.3 Related work

Factor analysis has been used extensively to analyze graded learner response data [19, 23]. While some factor analysis methods treat binary-valued graded learner responses as real numbers [2, 13], others use probabilistic models to achieve superior performance in predicting unobserved learner responses. These methods include the additive factor model (AFM) [9], instructional factors analysis (IFA) [10], and learning factor analysis (LFA) [22], which all assume that the number of concepts K to be known a priori. Collaborative filtering IRT (CF-IRT) [5] and SPARFA [17] both use cross-validation to select K, as well as all other tuning parameters, by identifying the best prediction performance on unobserved

learner responses. This approach is computationally extensive and does not scale to MOOC scale applications, where the dimension of the problem is large and immediate feedback is required. The authors of [4] proposed to select K by applying an SVD to the binary-valued graded learner response matrix and examining the decay of its singular values, which is not an automated approach.

Matrix completion (MC) aims to recover a low-rank matrix from incomplete, real-valued observations [6, 8], and has been used extensively in collaborative filtering applications. More recently, 1-bit MC [12], and its generalization, quantized MC [15] have been proposed for the recovery of low-rank matrices from incomplete binary-valued and quantized (or ordinal) observations, respectively. Since the graded learner responses in educational scenarios are typically binary-valued or ordinal, we next investigate the applicability of quantized MC [15] to educational scenarios.

2. SPARFA-LITE STATISTICAL MODEL

SPARFA-Lite aims at recovering the unknown, low-rank matrix $\mathbf{Z}$ that governs the learners' responses to questions, solely from quantized (ordinal) graded learner responses. Suppose that we have N learners answering Q questions. Let the $Q \times N$ matrix $\mathbf{Z}$ be the underlying low-rank matrix that we seek to recover. Let $Y_{i,j} \in \mathcal{O}$ denote the quantized observed graded response of the jth learner, with $j \in \{1, \ldots, N\}$, to the ith question, with $i \in \{1, \ldots, Q\}$. $\mathcal{O} = \{1, \ldots, P\}$ is a set of P ordered labels. Inspired by [15], we use the following model for the graded learner response $Y_{i,j}$:

$$Y_{i,j} = \mathcal{Q}(Z_{i,j} + \epsilon_{i,j}), \ (i,j) \in \Omega_{\text{obs}}, \quad (1)$$
$$\epsilon_{i,j} \sim Logistic(0,1).$$

Here, $Logistic(0,1)$ represents the Logistic distribution with zero mean and unit scale [14]. The set $\Omega_{\text{obs}} \subseteq \{1, \ldots, Q\} \times \{1, \ldots, N\}$ contains the indices associated to the observed learner responses $Y_{i,j}$. The function $\mathcal{Q}(\cdot) : \mathbb{R} \to \mathcal{O}$ represents a scalar quantizer, defined as

$$\mathcal{Q}(x) = p \quad \text{if } \omega_{p-1} < x \leq \omega_p, \ p \in \mathcal{O},$$

where $\{\omega_0, \ldots, \omega_P\}$ is a set of quantization bin boundaries, with $\omega_0 \leq \omega_1 \leq \cdots \leq \omega_{P-1} \leq \omega_P$. We will assume that the set of quantization bin boundaries $\{\omega_0, \ldots, \omega_P\}$ is known a priori. In situations where these bin boundaries are unknown, they can be estimated directly from data (see, e.g., [15, 16] for the details).

In terms of the likelihood of the observed graded learner responses $Y_{i,j}$, the model in (1) can be written equivalently as

$$p(Y_{i,j} = p \mid Z_{i,j}) = \Phi(\omega_p - Z_{i,j}) - \Phi(\omega_{p-1} - Z_{i,j}), \quad (2)$$

where $\Phi(x) = \frac{1}{1+e^{-x}}$ corresponds to the inverse logit link function. For this paper, we will be using only the inverse logit link function as it leads to algorithms with lower computational complexity comparing to the inverse probit link function [15].

The goal of the SPARFA-Lite algorithm detailed next is to recover the unknown low-rank matrix $\mathbf{Z}$ given the observed binary-valued graded learner responses $Y_{i,j}, (i,j) \in \Omega_{\text{obs}}$.

3. THE SPARFA-LITE ALGORITHM

To recover the low-rank matrix $\mathbf{Z}$ from binary-valued graded learner responses, we minimize the negative log-likelihood of the observed graded learner responses $Y_{i,j}, (i,j) \in \Omega_{\text{obs}}$, subject to a low-rank promoting constraint on $\mathbf{Z}$. In particular, we seek to solve

the following convex optimization problem:

$$(\text{P}) \begin{cases} \underset{\mathbf{Z} \in \mathbb{R}^{Q \times N}}{\text{minimize}} & f(\mathbf{Z}) = -\sum_{i,j:(i,j) \in \Omega_{\text{obs}}} \log p(Y_{i,j} | Z_{i,j}) \\ \text{subject to} & \|\mathbf{Z}\| \leq \lambda. \end{cases}$$

Here, the constraint $\|\mathbf{Z}\| \leq \lambda$ is used to promote low-rank solutions $\mathbf{Z}$ [8] and the parameter $\lambda > 0$ is used to control its rank. In practice, one can use the nuclear norm constraint $\|\mathbf{Z}\|_* \leq \lambda$, which is a convex relaxation of the (non-convex) low-rank constraint $\text{rank}(\mathbf{Z}) \leq r$ [6,8]; alternatively, one can use the max-norm constraint $\|\mathbf{Z}\|_{\text{max}} \leq \lambda$ (see [18] for the details). We select the only algorithm parameter $\lambda > 0$ via cross-validation. We emphasize that this parameter-selection process of SPARFA-Lite is much more efficient than that for regular SPARFA, which has three algorithm parameters.

Since the gradient of the negative log-likelihood of the inverse logit link function can be computed efficiently, (P) can be solved efficiently via the FISTA framework [3]. Starting with an initialization of the matrix $\mathbf{Z}$, at each inner iteration $\ell = 1, 2, \ldots$, the algorithm performs a gradient step that aims at reducing the objective function $f(\mathbf{Z})$, followed by a projection step that makes the solution satisfy the constraint $\|\mathbf{Z}\| \leq \lambda$. Both steps are repeated until convergence.

The *gradient step* is given by $\widehat{\mathbf{Z}}^{\ell+1} \leftarrow \mathbf{Z}^\ell - s_\ell \nabla f$, where s_ℓ is the step-size at iteration ℓ (see [15] for the details on step-size selection). The gradient of the objective function $f(\mathbf{Z})$ with respect to $\mathbf{Z}$ is given by

$$[\nabla f]_{i,j} = \begin{cases} \frac{\Phi'(L_{i,j} - Z_{i,j}) - \Phi'(U_{i,j} - Z_{i,j})}{\Phi(U_{i,j} - Z_{i,j}) - \Phi(L_{i,j} - Z_{i,j})} & \text{if } (i,j) \in \Omega_{\text{obs}} \\ 0 & \text{otherwise,} \end{cases}$$

where the derivative of the inverse logit link function corresponds to $\Phi'(x) = \frac{1}{2+e^{-x}+e^x}$. The $Q \times N$ matrices $\mathbf{U}$ and $\mathbf{L}$ contain the upper and lower bin boundaries corresponding to the measurements $Y_{i,j}$, i.e., we have $U_{i,j} = \omega_{Y_{i,j}}$ and $L_{i,j} = \omega_{Y_{i,j}-1}$.

The *projection step* imposes low-rankness on $\mathbf{Z}$. For the nuclear norm constraint case $\|\mathbf{Z}\|_* \leq \lambda$, this step requires a projection onto the nuclear norm ball with radius λ, which can be performed by first computing the SVD of $\mathbf{Z}$ followed by projecting the vector of singular values onto an ℓ_1-norm ball with radius λ (the details can be found in [6]). The resulting projection step corresponds to

$$\mathbf{Z}^{\ell+1} \leftarrow \widetilde{\mathbf{U}} \text{diag}(\mathbf{s}) \widetilde{\mathbf{V}}^T, \ \text{with} \ \mathbf{s} = \text{P}_\lambda(\text{diag}(\mathbf{S})), \quad (3)$$

where $\widetilde{\mathbf{U}} \mathbf{S} \widetilde{\mathbf{V}}^T$ denotes the SVD of $\widehat{\mathbf{Z}}^{\ell+1}$. The operator $\text{P}_\lambda(\cdot)$ denotes the projection of a vector onto the ℓ_1-norm ball with radius λ, which can be computed at low complexity [15]. For the max-norm constraint $\|\mathbf{Z}\|_{\text{max}} \leq \lambda$, the projection step can be calculated efficiently by following the method put forward in [18]. We emphasize that SPARFA-Lite is guaranteed to converge to a global optimum, since the problem (P) is convex.

4. SPARFA-LITE LEARNING ANALYTICS

We now demonstrate how SPARFA-Lite can be used to perform LA. To this end, we assume that there is tag information available for each question, i.e., there are a set of M user-defined labels (tags) associated with the Q questions, with each question associated with at least one tag. We define the $Q \times M$ question-tag matrix $\mathbf{T}$ with $T_{i,m} = 1$ if tag m is associated to question i, and $T_{i,m} = 0$ otherwise. We also define the $Q \times N$ matrix $\mathbf{A}$ with $A_{i,j} = \Phi(Z_{i,j}) \in [0,1]$, which is the de-noised and

completed version of the (partially observed) graded learner response matrix $\mathbf{Y}$. Using both matrices $\mathbf{T}$ and $\mathbf{A}$, we can compute the $N \times M$ learner tag knowledge matrix $\mathbf{B}$ with the entries $B_{j,m} = (\sum_{i=1}^{Q} T_{i,m})^{-1} \widetilde{B}_{j,m} \in [0,1]$, where $\widetilde{\mathbf{B}} = \mathbf{A}^T \mathbf{T}$. The entries $B_{j,m}$ represent the de-noised concept knowledge of learner j on tag m; large values represent good knowledge of tag m, whereas small values represent poor tag knowledge. This tag knowledge information is crucial for a PLS to perform LA.

5. EXPERIMENTS

We now compare SPARFA-Lite against existing factor analysis methods for predicting unobserved learner responses, using real-world educational datasets and demonstrate its efficacy in performing LA. All algorithm parameters are selected through cross-validation. All results are averaged over 25 independent Monte–Carlo trials.

5.1 Predicting unobserved learner responses

We first compare the performance of SPARFA-Lite in predicting unobserved graded learner responses with two state-of-the-art factor analysis algorithms.

Datasets. In this experiment, we use five different educational datasets for: (1) an undergraduate course on fundamentals of electrical engineering, consisting of $N = 92$ learners answering $Q = 203$ questions, with 99.5% of the answers observed; (2) an undergraduate course on signals and systems, consisting of $N = 41$ learners answering $Q = 143$ questions, with 97.1% of the answers observed; (3) an undergraduate course on introduction to probability and statistics, consisting of $N = 57$ learners answering $Q = 107$ questions, with 68.9% of the answers observed; (4) a university entrance exam, consisting of $N = 1706$ learners answering $Q = 60$ questions, with 60.9% of the answers observed; and (5) another university entrance exam, consisting of $N = 1564$ learners answering $Q = 60$ questions, with 70.8% of the answers observed. The undergraduate course datasets are collected via OpenStax Tutor [21]; see [25] for the details on the university entrance exam dataset. Note that all of these datasets contain binary-valued graded learner responses, which is a special case of the general, quantized model proposed above (with $P = 2$ and $\{\omega_0, \omega_1, \omega_2\} = \{-\infty, 0, \infty\}$). For simplicity, we will refer to the individual datasets as Dataset 1-to-5, respectively.

Experimental setup. We now compare SPARFA-Lite against CF-IRT [5] and SPARFA [17], two established factor analysis methods that perform well in terms of predicting unobserved graded learner responses. To assess prediction performance on unobserved learner responses, we randomly puncture each dataset by removing 20% of the observed learner responses in $\mathbf{Y}$ to form a test set. We then train all three algorithms on the rest of the observed learner responses and predict the unobserved responses in the test set. Since CF-IRT and SPARFA both have the number of concepts K as a tuning parameter, we run both algorithms using a range of possible values of K and select the value of K that achieves the best prediction performance. For SPARFA-Lite, we only need to select the value of the single algorithm parameter λ that controls K. To assess the prediction performance of all three algorithms, we use three well-established performance metrics: prediction accuracy (COR), prediction likelihood (LIK), and area under the receiver operation characteristic curve (AUC) [11]. The prediction accuracy corresponds to the percentage of correctly predicted responses. The prediction likelihood corresponds to the average the predicted likelihood of the unobserved responses, i.e.,

Table 1: Performance comparison of SPARFA-Lite vs. CF-IRT and SPARFA on predicting unobserved ratings for five educational datasets. Bold numbers represent the best performance among the three algorithms. SPARFA-Lite achieves comparable performance to CF-IRT and SPARFA in all experiments and metrics at significantly lower computational complexity.

		CF-IRT [5]	SPARFA [17]	**SPARFA-Lite**
	COR	0.8687	0.8711	**0.8737**
Dataset 1	LIK	**0.7286**	0.7195	0.7235
	AUC	0.8247	0.8056	**0.8299**
	COR	0.8061	0.8096	**0.8181**
Dataset 2	LIK	0.6393	**0.6759**	0.6707
	AUC	0.7985	0.7285	**0.8047**
	COR	**0.7263**	0.7000	0.7200
Dataset 3	LIK	**0.5876**	0.5334	0.5699
	AUC	**0.7629**	0.7116	0.7372
	COR	0.6967	0.7015	**0.7019**
Dataset 4	LIK	0.5538	**0.5587**	0.5537
	AUC	0.7180	**0.7249**	0.7175
	COR	0.6866	0.6880	**0.6903**
Dataset 5	LIK	0.5506	**0.5536**	0.5505
	AUC	0.7457	**0.7478**	0.7472

$\frac{\sum_{i,j:(i,j)\in\bar{\Omega}_{\mathrm{obs}}} p(Y_{i,j}|Z_{i,j})}{|\bar{\Omega}_{\mathrm{obs}}|}$, where $\bar{\Omega}_{\mathrm{obs}}$ represents the set of learner responses in the test set. The area under curve is a commonly-used performance metric for binary classifiers (see [11] for the details).

Results and discussion. Table 1 shows the mean of the performance metrics over 25 trials. We see that SPARFA-Lite achieves comparable performance as CF-IRT and SPARFA. Note that it outperforms CF-IRT and SPARFA on the most important performance metric–prediction accuracy (COR), with the exception of Dataset 3.

We emphasize that SPARFA-Lite is computationally more efficient than CF-IRT and SPARFA, since it (i) has only a single algorithm parameter and (ii) can be solved efficiently as it is a convex optimization problem. CF-IRT and SPARFA, in contrast, have multiple tuning parameters (including K) [5, 17], which means one have to run them multiple times to conduct a grid search over all possible values of these parameters. In particular, one Monte–Carlo trial of SPARFA-Lite on Dataset 1 only takes 3 sec, while CF-IRT and SPARFA require roughly 2 min. and 10 min. respectively, in MAT-LAB on a standard desktop PC with a 3.07 GHz Intel Core i7 processor (corresponding to $40\times$ and $200\times$ speed up). One can further reduce the computational complexity of SPARFA-Lite by replacing the nuclear norm constraint with the max-norm constraint [7, 18].

5.2 SPARFA-Lite learning analytics

Dataset and experimental setup. In this experiment, we use data collected from a high-school algebra test conducted on Amazon's Mechanical Turk [1]. The dataset consists of the quantized (with $P = 4$ ordinal values) graded responses of $N = 99$ learners answering $Q = 34$ questions, and the learner responses are fully observed. A total of $M = 13$ tags have manually been assigned to the questions. We use SPARFA-Lite to perform learning analytics on this dataset as described in Sec. 4.

Results and discussion. Table 2 shows the tag knowledge profile for a set of selected learners on the tags "Simplifying expressions," "Geometry," and "Systems of equations." The first row of

Table 2: Tag knowledge of selected learners. SPARFA-Lite performs robust LA by estimating each learner's tag knowledge from ordinal graded response data.

	Simplifying expressions	Geometry	System of equations
Class average	69 %	64%	30%
Best learner	84%	79%	34%
Average learner	70%	63%	24%
Worst learner	32%	34%	43%

the table shows the mean tag knowledge of all learners (in precent), while rows 2–4 show the tag knowledge (in percent) for the best learner, an average learner, and the worst learner, respectively. Leveraging these tag knowledge profiles, a PLS can automatically provide personalized feedback to learners on their strengths and weaknesses, and automatically recommend learning resources for remedial studies. For example, for the average learner in Table 2, a PLS would alert them to focus on the tag "System of equations" and recommend them learning resources associated with this tag, because their tag knowledge is below the class average. Moreover, a PLS can use this analysis to provide feedback to course instructors on the average tag knowledge of the entire class, helping them to make timely adjustments to their future course plan.

6. CONCLUSIONS

SPARFA-Lite is an efficient method that analyzes an incomplete set of quantized graded learner responses to questions to perform learning analytics. SPARFA–Lite achieves comparable or superior performance in predicting unobserved graded learner responses compared to existing factor-analysis methods, with significantly reduced computational complexity.

7. REFERENCES

[1] Amazon Mechanical Turk. http://www.mturk.com/mturk/welcome, Sep. 2012.

[2] T. Barnes. The Q-matrix method: Mining student response data for knowledge. In *Proc. American Association for Artificial Intelligence Workshop on Educational Data Mining*, July 2005.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Science*, 2(1):183–202, Mar. 2009.

[4] B. Beheshti, M. Desmarais, and R. Naceur. Methods to find the number of latent skills. In *Proc. 5th Intl. Conf. on EDM*, pages 81–86, June 2012.

[5] Y. Bergner, S. Droschler, G. Kortemeyer, S. Rayyan, D. Seaton, and D. Pritchard. Model-based collaborative filtering analysis of student response data: Machine-learning item response theory. In *Proc. 5th Intl. Conf. on EDM*, pages 95–102, June 2012.

[6] J. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optimization*, 20(4):1956–1982, Mar. 2010.

[7] T. Cai and W. Zhou. A max-norm constrained minimization approach to 1-bit matrix completion. *J. Machine Learning Research*, 14:3619–3647, Dec. 2014.

[8] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, Dec. 2009.

[9] H. Cen, K. R. Koedinger, and B. Junker. Learning factors analysis–a general method for cognitive model evaluation and improvement. In M. Ikeda, K. D. Ashley, and T. W. Chan, editors, *Intelligent Tutoring Systems*, volume 4053 of *Lecture Notes in Computer Science*, pages 164–175. Springer, June 2006.

[10] M. Chi, K. Koedinger, G. Gordon, and P. Jordan. Instructional factors analysis: A cognitive model for multiple instructional interventions. In *Proc. 4th Intl. Conf. on EDM*, pages 61–70, July 2011.

[11] 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, Dec. 1994.

[12] M. A. Davenport, Y. Plan, E. van den Berg, and M. Wootters. 1-bit matrix completion. *arXiv preprint:1209.3672*, Sep. 2012.

[13] M. Desmarais. Conditions for effectively deriving a Q-matrix from data with non-negative matrix factorization. In *Proc. 4th Intl. Conf. on EDM*, pages 41–50, July 2011.

[14] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2010.

[15] A. S. Lan, C. Studer, and R. G. Baraniuk. Matrix recovery from quantized and corrupted measurements. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, May. 2014, to appear.

[16] A. S. Lan, C. Studer, A. E. Waters, and R. G. Baraniuk. Tag-aware ordinal sparse factor analysis for learning and content analytics. In *Proc. 6th Intl. Conf. on Educational Data Mining*, pages 90–97, July 2013.

[17] A. S. Lan, A. E. Waters, C. Studer, and R. G. Baraniuk. Sparse factor analysis for learning and content analytics. *J. Machine Learning Research*, 2014, to appear.

[18] J. Lee, B. Recht, N. Srebro, J. Tropp, and R. Salakhutdinov. Practical large-scale optimization for max-norm regularization. *Advances in Neural Information Processing Systems*, 22:1297–1305, Dec. 2010.

[19] N. Li, W. W. Cohen, and K. R. Koedinger. A machine learning approach for automatic student model discovery. In *Proceedings of the 4th Intl. Conf. on Educational Data Mining*, pages 31–40, Jul. 2011.

[20] F. G. Martin. Will massive open online courses change how we teach? *Communications of the ACM*, 55(8):26–28, Aug. 2012.

[21] OpenStaxTutor. https://openstaxtutor.org/, Sep. 2013.

[22] P. I. Pavlik, H. Cen, and K. R. Koedinger. Learning factors transfer analysis: Using learning curve analysis to automatically generate domain models. In *Proc. 2nd Intl. Conf. on EDM*, pages 121–130, July 2009.

[23] J. C. Stamper, T. Barnes, and M. Croy. Extracting student models for intelligent tutoring systems. In *Proc. Natl. Conf. on Artificial Intelligence*, volume 22, pages 113–147, July 2007.

[24] K. VanLehn, C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill. The Andes physics tutoring system: Lessons learned. *Intl. J. of Artificial Intelligence in Education*, 15(3):147–204, Sep. 2005.

[25] D. Vats, C. Studer, A. S. Lan, L. Carin, and R. G. Baraniuk. Test size reduction for concept estimation. In *Proc. 6th Intl. Conf. on Educational Data Mining*, pages 292–295, July 2013.