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 ex-
tensively 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

2. SPARFA-LITE STATISTICAL MODEL

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

$$
Y_{i,j} = Q(Z_{i,j} + e_{i,j}), \quad (i,j) \in \Omega_{\text{obs}},
$$

$$
e_{i,j} \sim \text{Logistic}(0,1). \tag{1}
$$

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 $Q(\cdot) : \mathbb{R} \to O$ represents a scalar
quantizer, defined as

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

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 re-
sponses $Y_{i,j}$, the model in (1) can be written equivalently as

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

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 com-
plexity 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 $Z$ 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 $Z$ 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 $Z$. In particular, we seek to solve
the following convex optimization problem:

$$
(P) \quad \text{minimize} \quad f(Z) = - \sum_{i,j:(i,j) \in \Omega_{\text{obs}}} \log p(Y_{i,j} | Z_{i,j})
$$

subject to $\|Z\| \leq \lambda$.

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

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

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

$$
\nabla f|_{Z^{\ell}} = \begin{cases}
\Phi'(U_{i,j} - Z_{i,j}) - \Phi'(U_{i,j} - Z_{i,j}) & \text{if } (i,j) \in \Omega_{\text{obs}},
0 & \text{otherwise},
\end{cases}
$$

where the derivative of the inverse logit link function corresponds
to $\Phi'(x) = \frac{1}{1+e^{-x}}$. The $Q \times N$ matrices $U$ and $L$ 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 $Z$. For the nuclear
norm case $\|Z\|_* \leq \lambda$, this step requires a projection onto
the nuclear norm ball with radius $\lambda$, which can be performed
by first computing the SVD of $Z$ followed by projecting the vector
of singular values onto an $\ell_1$-norm ball with radius $\lambda$ (the
details can be found in [6]). The resulting projection step corresponds to

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

where $\mathbf{U} \mathbf{S} \mathbf{V}^T$ denotes the SVD of $Z^{\ell+1}$. The operator $\mathbf{P}_\lambda(\cdot)$
denotes the projection of a vector onto the $\ell_1$-norm ball with radius $\lambda$,
which can be computed at low complexity [15]. For the max-norm
constraint $\|Z\|_{\text{max}} \leq \lambda$, the projection step can be calculated effi-
ciently 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 asso-
ciated with at least one tag. We define the $Q \times M$ question-
tag matrix $T$ 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
$A$ 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 $Y$. Using both matrices $T$ and $A$, we can compute the $N \times M$ learner tag knowledge matrix $B$ with the entries $B_{m,j} = (\sum_{i=1}^{N} T_{i,m})^{-1} B_{i,m} \in [0, 1]$, where $\bar{B} = AT$. The entries $B_{m,j}$ 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 99.95% 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 $\{\omega_1, \omega_2\} = \{-\infty, 0, \infty\}$). 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 answers observed; (2) an under-

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
the table shows the mean tag knowledge of all learners (in percent), 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