Personalizing Embedded Assessment Sequences in Narrative-Centered Learning Environments: A Collaborative Filtering Approach

Wookhee Min, Jonathan P. Rowe, Bradford W. Mott, and James C. Lester

Department of Computer Science, North Carolina State University, Raleigh, NC 27695
{wmin, jprowe, bwmott, lester}@ncsu.edu

Abstract. A key challenge posed by narrative-centered learning environments is dynamically tailoring story events to individual students. This paper investigates techniques for sequencing story-centric embedded assessments—a particular type of story event that simultaneously evaluates a student's knowledge and advances an interactive narrative's plot—in narrative-centered learning environments. We present an approach for personalizing embedded assessment sequences that is based on collaborative filtering. We examine personalized event sequencing in an edition of the CRYSTAL ISLAND narrative-centered learning environment for literacy education. Using data from a multi-week classroom study with 850 students, we compare two model-based collaborative filtering methods, including probabilistic principal component analysis (PPCA) and non-negative matrix factorization (NMF), to a memory-based baseline model, *k*-nearest neighbor. Results suggest that PPCA provides the most accurate predictions on average, but NMF provides a better balance between accuracy and run-time efficiency for predicting student performance on story-centric embedded assessment sequences.

Keywords: Narrative-Centered Learning Environments, Embedded Assessment, Collaborative Filtering.

1 Introduction

Over the past several years there has been growing interest in narrative-centered learning environments, a class of game-based learning environments that contextualize learning and problem solving within interactive story scenarios. A key benefit of narrative-centered learning environments is their capacity to discreetly support students' learning processes by tightly integrating instructional and narrative elements. By leveraging the motivational characteristics of narrative and games, along with the adaptive pedagogy of intelligent tutoring systems, narrative-centered learning environments create educational experiences that are situated in meaningful contexts, found to be highly engaging, and dynamically personalized to individual students. Narrative-centered learning environments are under investigation in a range of domains, such as language learning [1], anti-bullying education [2], intercultural negotiation training [3], middle school science [4], and network security [5].

A key challenge posed by narrative-centered learning environments is how to dynamically tailor story events to individual students [5–7]. Events in narrative-centered learning environments fulfill dual roles: they advance emerging plots in which students are active participants, and they serve pedagogical purposes such as providing feedback or assessments. In order to satisfy these dual roles, events take many forms. For example, a student found to have a misconception may be prompted to complete a quest that will help remediate his knowledge, or a student may find a virtual book accompanied by half-completed notes to be filled out, a form of embedded assessment. Because students often have considerable autonomy in narrative-centered learning environments, students may trigger events in many possible orders, including sequences that are sub-optimal for learning or engagement. In order to cope with this uncertainty, narrative-centered learning environments must be capable of dynamically personalizing event sequences to preserve the environment's ability to satisfy instructional and narrative objectives.

This paper investigates a method for personalizing event sequences in narrative-centered learning environments based on collaborative filtering. Frequently used in recommender systems, collaborative filtering techniques make predictions about individuals' actions based on the actions of others who behave similarly. We focus on personalizing a particular class of story events, story-centric embedded assessments, in the CRYSTAL ISLAND narrative-centered learning environment. CRYSTAL ISLAND features a mystery scenario about a spreading outbreak, and it has recently been extended to incorporate a curricular focus of middle-grade literacy education. Story-centric embedded assessments in CRYSTAL ISLAND evaluate students' reading comprehension skills, with a focus on complex informational texts and concept matrices, while simultaneously providing clues for solving the mystery. We examine several collaborative filtering algorithms for personalizing embedded assessment sequences. We compare two model-based collaborative filtering algorithms, probabilistic principal component analysis (PPCA) and non-negative matrix factorization (NMF), to a memory-based baseline model, *k*-nearest neighbor (kNN). Results suggest that PPCA provides the most accurate predictions on average, but NMF provides a better balance between accuracy and run-time efficiency.

2 Related Work

Narrative-centered learning environments couple salient features of stories (rich settings, believable characters, and compelling plots) and digital game environments (interactivity, rewards, and feedback) in order to increase student motivation, support meaning making, and guide complex problem solving. Interactions with narrative-centered learning environments can take several forms. Students may directly influence a narrative by completing actions in order to solve a problem [4, 5, 8], or they may indirectly influence events by providing guidance to autonomous virtual characters [2]. Multi-user virtual environments such as River City [4] use rich narrative settings to contextualize inquiry-based science learning scenarios with social and collaborative elements. Other work has utilized interactive narrative generation and agent behavior planning to create adaptive narrative experiences [1– 2].

Narrative-centered learning environments have begun to directly incorporate intelligent tutoring facilities, which provide support for coaching, feedback, and reflection that is tailored to individual students [1, 3, 5, 6, 7]. Many of these systems formalize models for personalizing story event sequences in terms of rule-based techniques, STRIPS-style planning algorithms, or probabilistic graphical models. In many cases, past approaches have involved hand-authoring problem domains [4–6], a process that can be labor-intensive, or supervised machine learning techniques that require training data collected in laboratory settings, such as Wizard of Oz experiments [7]. Our work is the first to use collaborative filtering for tailoring story-centric embedded assessments in narrative-centered learning environments, and the models are induced directly from student interaction data collected in classroom settings. Our approach is inspired by recent work on collaborative filtering-based drama management by Yu and Riedl [9].

Figure 1. CRYSTAL ISLAND narrative-centered learning environment.

3 CRYSTAL ISLAND

Over the past several years, our lab has been developing CRYSTAL ISLAND (Figure 1), a narrative-centered learning environment for middle school microbiology [8]. Designed as a supplement to classroom science instruction, CRYSTAL ISLAND's curricular focus has been expanded to include literacy education based on Common Core State Standards for reading informational texts. The narrative focuses on a mysterious illness afflicting a research team on a remote island. Students play the role of a visitor who is drawn into a mission to save the team from the outbreak. Students explore the research camp from a first-person viewpoint, gather information about patient symptoms and relevant diseases, form hypotheses about the infection and transmission source, use virtual lab equipment and a diagnosis worksheet to record their findings, and report their conclusions to the camp's nurse.

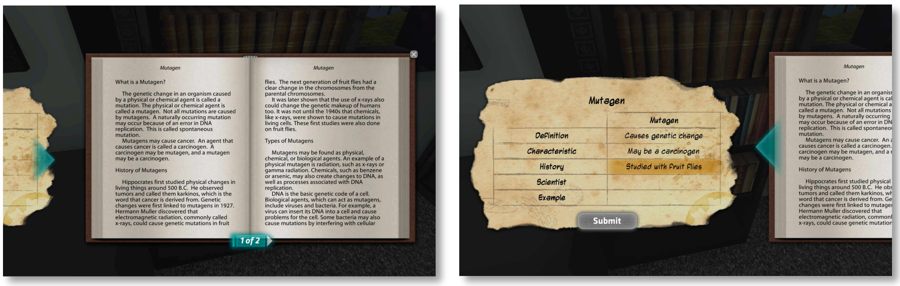

Figure 2. (Left) An informational text stylistically formatted like a virtual book, and (Right) a concept matrix stylistically formatted as a scrap of note paper.

As part of CRYSTAL ISLAND's curricular focus on literacy, students encounter books and articles throughout the camp that contain complex informational texts about microbiology concepts (Figure 2, left). Students read and analyze these texts, as well as complete associated concept matrices, to acquire knowledge necessary to diagnose the illness. Concept matrices (Figure 2, right) are framed within the narrative as partially completed notes written by one of the research team's scientists. Students learn that the scientist has fallen ill, and they must now "complete" the notes based on content in the informational texts. The concept matrices are story-centric embedded assessments that evaluate students' reading comprehension skills by requiring students to recognize and make connections among key ideas from the informational texts. Completing a concept matrix involves making several selections to populate blank cells based on the adjacent informational text.

Within the CRYSTAL ISLAND narrative environment, virtual books and articles have fixed physical positions. Because narrative-centered learning environments such as CRYSTAL ISLAND support many possible problem-solving paths, students encounter objects in many different orders. If the content of each book and article is static, students may encounter embedded assessments in orders that are sub-optimal for learning or solving the mystery. Instead, when a student opens a book or research article in the virtual environment, the informational text content should be dynamically selected and personalized to the student in terms of subject and difficulty level. In order to meet this objective, we have designed CRYSTAL ISLAND to draw on a pool of 27 informational texts and concept matrices that can be arbitrarily assigned as the contents of books and articles during run-time. The method that we use to personalize embedded assessment sequences is collaborative filtering.

4 Collaborative Filtering for Sequencing Embedded Assessments

Popularized by their use in recommender systems, collaborative filtering (CF) algorithms are used to predict user preferences about unseen items using ratings from similar users. The underlying assumption of CF is that if multiple users have similar past interests, they will also have similar preferences for items they have not yet encountered [10–11]. We investigate two model-based CF methods for personalizing story-centric embedded assessment sequences in the CRYSTAL ISLAND narrative-centered learning environment. The first is model-based techniques, which are particularly useful in domains with data sparsity, an inherent issue in sequencing story-centric embedded assessments. Similar to prefix-based collaborative filtering [9], our work focuses on recommending entire sequences of embedded assessments, rather than recommendations of individual assessments. Because the number of possible assessment sequences grows exponentially with sequence length, students will never experience the vast majority of assessment sequences, even in cases of large training data sets. In order to evaluate model-based CF techniques' ability to cope with the resulting data sparsity, we examine two dimension-reduction methods for personalizing story-centric embedded assessments: non-negative matrix factorization (NMF) and probabilistic principal component analysis (PPCA). In conjunction with each of NMF and PPCA, we employ expectation maximization (EM); through its iterative maximum likelihood estimation process, EM replaces missing values in the data set along with NMF and PPCA [12].

4.1 Non-negative Matrix Factorization

NMF is a decomposition technique for an observed matrix R populated by multivariate data. NMF involves finding two matrix factors, typically with reduced dimensionality relative to R, that approximate R by their multiplication [13].

$$R \approx W \times H . \quad (1)$$

The NMF algorithm is represented by Equation 1, where R is an $n \times m$ matrix denoting the observed input data, n is the number of different assessment sequences, and m is the number of observed users. The $n \times r$ matrix W is a basis model, which is calculated through the NMF algorithm, and the $r \times m$ matrix H is a coefficient matrix based on the basis model W.

NMF requires that two constraints be met by the input data matrix: (1) all values in R, W, and H are non-negative, and (2) r is smaller than either m or n. Because W is typically smaller in size than R due to the second constraint, a key aspect of the NMF algorithm lies in how W encodes the hidden structure in the matrix R. This is performed through an iterative process seeking the maximum likelihood estimate of the model's parameters [13–14].

4.2 Probabilistic Principal Component Analysis

The PPCA algorithm is a probabilistic extension of traditional principal component analysis, which finds the principal axes of a set of observed data vectors through iterative maximum likelihood estimation by the EM algorithm [15–16].

$$t = Wx + \mu + \varepsilon . \quad (2)$$

The PPCA technique is represented by Equation 2, where t denotes a d-dimensional observation vector; in our case, this is comprised of a single student's in-game assessment scores (Equation 3). The $d \times q$ matrix W contains principal components in its columns, x refers to a q-dimensional latent variable that is related to the observed data t by W, μ is a non-zero mean value, and ε is a Gaussian noise parameter. Because ε is normally distributed (i.e., it follows $N(0, \sigma^2 I)$), the distribution of the observation vector t given the latent variable x can be represented as $t|x \sim N(Wx + \mu, \sigma^2 I)$. Since the marginal distribution over the latent variable x follows a Gaussian distribution, the marginal distribution over t is also Gaussian [15].

$$Assessment\ Score = \frac{Number\ of\ assessments\ solved\ correctly\ so\ far}{Number\ of\ assessments\ attempted\ so\ far} * 100.0 \tag{3}$$

5 Empirical Evaluation

To evaluate the collaborative filtering approach for tailoring story-centric embedded assessment sequences, we analyzed student interaction data from a teacher-led deployment of CRYSTAL ISLAND in two rural school districts. Students used CRYSTAL ISLAND over several weeks in their Language Arts classrooms. CRYSTAL ISLAND was an instructional anchor in a curricular unit on reading comprehension, which included supplementary learning activities. Prior to beginning the unit, and immediately following the unit, students completed web-based pre- and post-study assessments to measure their reading comprehension skills. Each assessment was comprised of three distinct pairs of informational texts and concept matrices; they mirrored the embedded assessments in CRYSTAL ISLAND, but covered different microbiology topics and did not include the stylistic appearance of assessments in the narrative environment.

The data set for our analysis included interaction logs and pre/post measures for 850 students. There were 436 males and 414 females. On average, students played CRYSTAL ISLAND for approximately 92 minutes over several class periods, they attempted 9.9 embedded assessments, they correctly filled out 7.2 concept matrices, and they failed to complete 1.7 concept matrices (i.e., after three incorrect attempts, the student was given the correct answers and prompted to move on).

Prior to investigating collaborative filtering techniques for personalizing story-centric embedded assessment sequences, we investigated whether students achieved significant improvements in their reading comprehension skills as a result of the CRYSTAL ISLAND unit. Matched pairs t-tests comparing pre-study (M=0.60, SD=0.26) to post-study (M=0.74, SD=0.28) assessment scores indicated that students' learning gains were statistically significant, $t(863) = 16.21$, $p < .01$. Next, we investigated whether students' use of CRYSTAL ISLAND impacted their gains in reading comprehension skills. Table 1 presents findings from a multiple regression analysis, which treated pre-study assessment score and average in-game assessment score as predictor variables, and post-study assessment score as a dependent variable. The regression model explained a significant proportion of the variance in post-study assessment scores, *Adj.* $R^2 = .332$. Results indicate that students' in-game assessment performances predict their improvements across pre- and post-study assessments.

These findings provide the foundation for investigating personalized sequencing of embedded assessments in CRYSTAL ISLAND. Our examination of collaborative filtering consists of three stages. First, we transformed the interaction log data into a format suitable for collaborative filtering analysis. This involves extracting raw

Table 1. Multiple Regression Analysis Predicting Post-Study Assessment Scores.

Dependent Variables	Independent Variables	*B*	*SE(B)*	β	*t*	*p*
Post-Study Assessment	Pre-Study Assessment	.511	.033	.476	15.620	< .001
	In-Game Assessment Score	.003	.000	.191	6.289	< .001

interaction logs from a MySQL database, filtering observations of students' in-game assessment performances, and constructing an assessment sequence matrix. The assessment sequence matrix, R, is comprised of n rows, one for each distinct sequence of embedded assessments, and m columns, one for each student. Each value in R is a student's in-game assessment score. Assessment scores range from 0 to 100. Any sequence not encountered by a student corresponds to a missing value in the matrix R.

Second, we reduced the sequencing task's dimensionality. The task's dimensionality grows exponentially with sequence length. Consequently, we reduced the number of considered sequences by focusing on a subset of the 27 informational texts in CRYSTAL ISLAND. We employed a Chi-square selection algorithm to choose five informational text/concept matrix pairs that were significantly correlated with solving the mystery. The five pairs consisted of the following topics: *Investigating an Illness*, *Salmonellosis*, *Microbes*, *Carcinogens*, and *Viruses*. Furthermore, while students could encounter story-centric embedded assessments in almost any order, the *Investigating the Illness* text was the first assessment for 735 of the students. In order to further reduce data sparsity, we considered only the 735 students who encountered the *Investigating an Illness* text as their first embedded assessment. This reduction produced an observation data matrix with 65 rows and 735 columns. Even with the steps taken to address data sparsity, 94.2% of the matrix values remained missing.

Third, we examined non-negative matrix factorization (NMF) and probabilistic principal component analysis (PPCA) across a range of reduced dimension values using 10-fold cross validation. Generating models for training involved the following: (1) replacing missing values for each student with the mean value of the student's non-missing assessment scores, (2) applying a collaborative filtering technique to the assessment sequence matrix, and (3) updating the missing values with new estimates if the distance between the original matrix and the new model-predicted matrix is decreased. Steps (2) and (3) are repeated until the distance is less than a threshold [14]. The validation step works similarly, except it uses a matrix inversion method instead of collaborative filtering, since it uses the previously trained model to predict scores for the validation set. Results from the evaluation are presented next.

6 Results

In order to evaluate the models' performance, we investigated their ability to minimize root mean square error (RMSE), which is defined in Equation 4.

$$RMSE = \sqrt{\frac{1}{|O|}\sum_{i,j\in O}(R^{v}_{i,j} - R^{v\prime}_{i,j})^2}\ . \qquad (4)$$

In the above equation, O denotes all coordinates for observed values (i.e., non-missing values), $|O|$ is the number of observed coordinates, $R^{v}_{i,j}$ is the value at i^{th} row and j^{th} column in the validation set, and $R^{v\prime}_{i,j}$ is the estimated value at i^{th} row and j^{th} column according to the inference mechanism based on a generated model.

In addition to NMF and PPCA, we implemented a memory-based approach to collaborative filtering, k-nearest neighbor (kNN), as a non-trivial baseline. In implementing kNN, we utilized Euclidean distance as a distance metric, inferring users' ratings by averaging assessment performance scores among the k "nearest" users. Similar to the examinations of NMF and PPCA, we investigated the performance of kNN across a range of k values using 10-fold cross validation.

The RMSE results for each model are presented in Figure 3. RMSE values are displayed on the y-axis; lower RMSE values are better. The two model-based collaborative filtering approaches substantially outperformed kNN. PPCA yielded the lowest average *RMSE* (M = 0.42), followed by NMF (M = 0.89) and kNN (M = 14.93). NMF achieved its best performance with 50 dimensions, and PPCA achieved comparably good performance at 55 dimensions, with *RMSE* = 0.18.

Given the real-time performance requirements of narrative-centered learning environments, we also elected to investigate the running times of each algorithm across 10-fold cross validation. The results of this comparison are shown in Figure 4. Due to the wide range in run-times among the three algorithms, we charted run-time on a base-2 logarithmic scale along the y-axis. PPCA (M=1998179ms) proved to be considerably slower than NMF (M=1056ms), raising concerns about its utility for run-time settings. Possible methods for improving the run-time performance of PPCA include adjusting its termination threshold value, or restricting the maximum

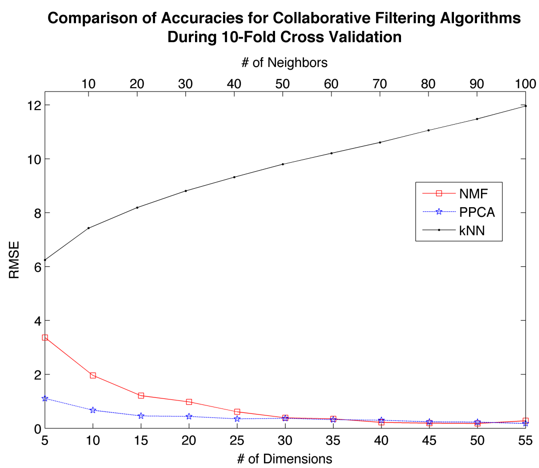

Figure 3. RMSE values for the three CF algorithms plotted across parameter values. The bottom x-axis displays dimension values for NMF and PPCA. The upper x-axis displays k values for kNN.

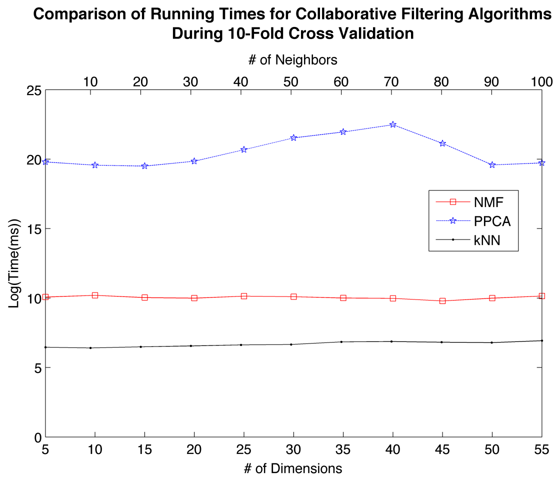

Figure 4. Run-times (ms) for the three CF algorithms during cross validation. Run-times are plotted on a logarithmic scale along the y-axis. The test machine utilized an Intel i7-2600K processor and 16GB RAM.

number of iterations used during maximum likelihood estimation. In the current setting, the termination threshold value was set to 0.001 for both NMF and PPCA, and no limit was specified on the maximum number of iterations.

7 Conclusions and Future Work

Collaborative filtering algorithms show considerable promise for dynamically tailoring story events in narrative-centered learning environments. This paper introduces an effective approach to personalizing sequences of story-centric embedded assessments. Using data from classroom studies of a literacy-focused edition of the CRYSTAL ISLAND narrative-centered learning environment, an empirical evaluation demonstrated that model-based collaborative filtering techniques, such as non-negative matrix factorization and probabilistic principal component analysis, outperform baseline approaches for accurately predicting student performance on embedded assessments of reading comprehension skills. Further, results suggest that NMF techniques provide a superior balance between predictive accuracy and running time compared to PPCA. In the future, we intend to investigate alternate collaborative filtering techniques with improved scalability for larger data sets and longer assessment sequences. Additionally, we will incorporate collaborative filtering models into CRYSTAL ISLAND to evaluate the impacts of tailoring story-centric embedded assessment sequences in run-time narrative-centered learning environments.

Acknowledgments. The authors wish to thank colleagues from the IntelliMedia Group and East Carolina University for their assistance. This research was supported by the National Science Foundation under Grant DRL-0822200. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Additional support was provided by the Bill and Melinda Gates Foundation, the William and Flora Hewlett Foundation, and EDUCAUSE.

References

1. Johnson, W.L.: Serious use of a Serious Game for Language Learning. In: 13th International Conference on Artificial Intelligence in Education, pp. 67-74 (2007)
2. Aylett, R., Louchart, S.: If I were you: double appraisal in affective agents. In: 7th international joint conference on Autonomous agents and multiagent systems, pp. 1233-1236 (2008)
3. Kim, Hill, Durlach, Lane, Forbell, Core, Marsella, Pynadath, Hart.: BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. Int. J. AIED, 19, pp. 289-308 (2009)
4. Nelson, B. C., Ketelhut, D. J.: Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. International Journal of Computer-Supported Collaborative Learning, 3(4), pp. 413-427 (2008)
5. Thomas, J. M., Young, R. M.: Annie: Automated Generation of Adaptive Learner Guidance for Fun Serious Games. IEEE Transactions on Learning Technologies, 3(4), pp. 329-343 (2010)
6. Mott, B., Lester, J.: Narrative-centered tutorial planning for inquiry-based learning environments. In: 8th International Conference on Intelligent Tutoring Systems, pp. 675–684 (2006)
7. Lee, S., Mott, B., Lester, J.: Real-time narrative-centered tutorial planning for story-based learning. In: 11th International Conference on Intelligent Tutoring Systems, pp. 476–481 (2012)
8. Rowe, J. P., Shores, L. R., Mott, B. W., Lester, J. C.: Integrating learning, problem solving, and engagement in narrative-centered learning environments. International Journal of Artificial Intelligence in Education, 21(2), pp. 115-133 (2011)
9. Yu, H., Riedl, M. O.: A sequential recommendation approach for interactive personalized story generation. In: 11th International Conference on Autonomous Agents and Multiagent Systems, pp. 71-78 (2012)
10. Resnick P., Varian H. R.: "Recommender systems," Communications of the ACM, 40(3), pp. 56–58 (1997)
11. Goldberg, K., Roeder, T., Gupta, D., Perkins, C.: Eigentaste: A Constant Time Collaborative Filtering Algorithm, Information Retrieval, 4(2), pp. 133-151 (2001)
12. Rubin, D. B., Thayler D. T.: EM algorithms for ML factor analysis, Psychometrika, 47(1), 69-76 (1982)
13. Lee, D. D., Seung, H. S.: Algorithms for non-negative matrix factorization, In: Advances in Neural Information Processing, pp. 556-562 (2000)
14. Zhang, S., Wang, W., Ford, J., Makedon, F.: Learning from incomplete ratings using non-negative matrix factorization, In: 6th SIAM Conference on Data Mining, 549-553 (2006)
15. Tipping, M. E., Bishop, C. M.: Probabilistic principal component analysis, Journal of the Royal Statistical Society, Series B, 61, 611-622 (1999)
16. Roweis, S.: EM Algorithms for PCA and SPCA, In: Advances in Neural Information Processing Systems 10, pp. 626-632 (1998)