Teacher Usage Behaviors within an Online Open Educational Resource Repository

Jennifer Sabourin
SAS Institute, Inc.
Cary, NC USA
Jennifer.Sabourin@sas.com
+1 919 531 3313

Lucy Kosturko
SAS Institute, Inc.
Cary, NC USA
Lucy.Kosturko@sas.com
+1 919 531 3430

Scott McQuiggan
SAS Institute, Inc.
Cary, NC USA
Scott.McQuiggan@sas.com
+1 919 531 1119

ABSTRACT
With instructional methods such as MOOCs and flipped classrooms rapidly gaining popularity and school budget cuts becoming more prevalent across the nation, increasing the usability of Open Educational Resources (OER) is highly relevant for today’s educators. Although several OER databases exist providing access to hundreds of thousands of resources, navigating these spaces, evaluating resources, and integrating them within classroom instruction has proven less than efficient. The present research explores learning analytics for understanding real-world interaction patterns with SAS® Curriculum Pathways®, which has over 120,000 active teacher users and over 1,300 freely available resources across multiple disciplines. In this preliminary investigation, users are clustered based on overall usage patterns. Patterns of resource interaction are then identified using association analysis. Results of this exploratory investigation provide insight into how users interact with large OER databases and introduce many avenues for continued investigation.

Author Keywords
Open educational resources; recommendation systems

INTRODUCTION
Open Educational Resource is a broad term that describes any “teaching, learning, and research resource that resides in the public domain or have been released under an intellectual property license that permits their free use or repurposing by others” [1]. Presently, hundreds of thousands of OER are available for K-12 use. While the abundance of OER is a good problem to have, individual instructors are spending a great deal of time tirelessly searching through this corpus to find tools that fit their specific needs [1]. In fact, some districts have reported million-dollar budgets allocated to the creation of OER repositories to support educators in their search [2]. Other efforts, such as Khan Academy (khanacademy.org), the Learning Registry (learningregistry.org), and OpenEd (openEd.io), have made finding and integrating OER more efficient with carefully designed user interfaces and useful filtering options. However, simply finding resources is only a portion of an educator’s OER integration process; evaluating and devising an integration plan adds additional complexities.

Understanding how users interact with and navigate a large repository of unique resources is a broad area of investigation. This exploratory analysis represents a preliminary step in this line of work by first empirically identifying groups of users and then examining differential patterns of use among them. Specifically, we are interested in how users find resources of interest, which resources are commonly used together, and if any of these patterns can be associated with success in navigating the repository. By exploring this line of work we hope to build more robust systems for helping educators find the set of resources that are most aligned to their goals.

PRELIMINARY INVESTIGATION
For the purpose of this preliminary investigation we consider user interaction data with SAS Curriculum Pathways from the past 5 years (August 2008 - August 2013). Available at no cost, SAS Curriculum Pathways provides interactive, standards-based resources in the core disciplines (English Language Arts, Mathematics, Science, Social Studies, and Spanish) for traditional, virtual, and home schools. At present, more than 120,000 users in more than 45,000 schools are actively using SAS Curriculum Pathways as part of their classroom instruction. The corpus used for this analysis focuses on data from approximately 87,000 users interacting with 1,524 unique resources. In total this corpus represents 4.9 million interactive sessions and observations from over 10.4 million resource hits. The majority of users (66.3%) identify themselves as educators, while 22.0% identify themselves as students, and the remaining do not list their role.

Method
The exploratory analyses include two phases: clustering of users and association analysis of resources use. In the first phase groups demonstrating similar patterns of usage are identified. These groups are then used in the second phase to further explore how user groups interact with resources and navigate the repository.
Clustering

K-means clustering was used to identify meaningful groups of users. The variables of interest in creating the clusters were: total sessions, total resources hits, average duration of time between sessions, and proportion of [English, Science, Social Studies, Mathematics and unclassified] resources. Each of these variables was heavily skewed so the log of the values were used as input to the clustering algorithm. K-means clustering was run with k=3, 4, and 5 and the clusters with greatest distance between means were selected (k=3). Initial exploration of the clusters indicated that clusters were selected based on their overall usage patterns.

Results showed that users in Cluster 1 (N = 3,869) demonstrated the highest usage. They logged in more frequently (M = 1180.5, SD = 8690.1) and interacted with a greater total number of resources (M = 2342.1, SD=14877.5). Based on these patterns, we consider this group Established Users. They show sustained interaction with the tools over time and, though there are fewer of these users, they make up a large proportion of daily hits to the repository. Cluster 2 (N = 61,411) demonstrated the lowest usage. Most of these users logged in only a few times (M=1.6, SD = 1.1) after creating an account. They looked at a few resources (M = 5.3, SD = 5.1) and then often did not return. We name this group Glancing Users and consider this group a target for future analysis and intervention as they seem to have the most difficulty in navigating the repository to find resources of interest. Finally, Cluster 3 (N = 21,589) demonstrated an intermediate pattern. During the timespan of the corpus, they logged in an average of approximately 14.4 (SD = 34.7) times and interacted with 50.7 (SD = 56.1) resources total. We name this group Browsing Users as their behaviors indicate that they are exploring the repository to identify resources to use again.

Association Analysis

The next set of analyses looked at specific patterns of resource use. Specifically, we were interested in whether users used particular resources together and how they found new resources. An association analysis was run independently for each of the three groups. This exploratory analysis produces a list of rules linking resources that are commonly used together by the users.

Results from the Glancing Users indicate that there are few common patterns in exploring resources. The only connections occur between some of the most commonly used resources in SAS Curriculum Pathways. Perhaps these resources were accessed based on a recommendation of a peer but then the user never explored the remaining resources, as is suggested by the total number of resources used by this group. Alternatively, patterns from the Browsing Users support the hypothesis that these users are exploring the space of resources. There is a spreading pattern from the most common resources out to many other resources supported by a long network of connections.

These users also utilize more resources per session (M = 5.4, SD = 6.2) than the Glancing Users (M =3.6, SD = 3.7) or even the Established Users (M = 3.2, SD = 4.0) further supporting the idea that these users are exploring the resources available in the repository. The Established Users on the other hand show less connectedness than Browsing Users, and instead favor resource clusters that are based on a particular domain. These experienced users have likely identified the set of resources that work best for their purposes. They may return to them often and not necessarily explore other options.

DISCUSSION

This preliminary investigation used k-means clustering and association analysis to understand the navigation patterns of users in SAS Curriculum Pathways. Analyses of a large corpus of interaction data indicated three groups of users: Glancing Users, Browsing Users, and Established Users, each with their own profile of resource navigation. As a preliminary analysis, this work pointed to many areas for future investigation.

One line of work will be investigating what characteristics and patterns of behavior differentiate Glancing and Browsing Users when they first begin interacting with the repository. Specifically, how can we encourage or facilitate better browsing behavior to increase the likelihood that users will find resources of interest and continue using the repository. Another key question is whether Browsing Users eventually transition to Established Users and how can this process be facilitated? Specifically, it will be important to investigate changes in patterns of behavior across time and to explore how users identify key resources that they would like to continue to use. It is possible that these resources may be the keystone resources identified in the association graphs. These analyses will require expanding the analyses beyond 5 years to include a user’s entire span of interaction. By exploring these patterns temporally situated for each user we may be able to better predict which users are most likely to return to the system, use it effectively and why these differences occur. Finally, it will be important to explore the search patterns of user groups when they first enter the repository. Do they use fundamentally different search keywords? How do they select resources after a search? This information can guide the resources that are displayed based on a search and may improve similar recommendations. Overall, it is hoped that this line of resource can further the understanding of educators’ navigation and use of large OER repositories.

REFERENCES