



## **An Associative Based Approach to Analyzing an Online Learning Environment**

**Ms. Bahareh Azarnoush, Arizona State University**

BAHAREH AZARNOUSH is a PhD student at the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, PO Box 878809, Tempe, AZ 85287-8809, bazarnou@asu.edu. Her research interest is statistical learning.

**Dr. Jennifer M Bekki, Arizona State University**

**Dr. Bianca L. Bernstein, Arizona State University**

**Dr. George C Runger, Arizona State University**

# An Associative Based Approach to Analyzing an Online Learning Environment

## Abstract

Recent years have shown an increase in both in the number and use of online educational learning environments. Correspondingly, there is a greater availability of rich data sets that describe both the learners themselves and their interactions with the online learning environment. In this paper, we demonstrate the use of a data mining tool, association analysis, to analyze this data. We demonstrate its applicability in understanding how learners use a particular online learning environment and for the identification of learner interactions with the environments that are associated with particular learning outcomes. The methodology is first described and then is demonstrated as a case study through its application to the *CareerWISE* online learning environment.

## Introduction

Recent years have witnessed a tremendous proliferation of the World Wide Web. This has changed the way we experience many aspects of everyday life, including the ways we communicate, conduct business, and educate. The advent of “online education” can be linked back to the mid-1970s when educators started adopting e-mail and online conferencing<sup>1</sup>. Since then, great advances in education delivery through the Web have been made, which have, in turn, overcome some of the geographical and temporal boundaries to education that previously existed.

With the increase in internet usage has come rich data sources for analyses<sup>2,3</sup> in applications such as e-commerce<sup>4</sup> and education<sup>5</sup>. While much of the work on analyzing this data has focused on e-commerce, there has been a growing trend toward the application of web-based data analysis for improving education. For example, past studies have demonstrated the use and analysis of data gathered from online learning environments (OLEs) to better understand learner usage of OLEs<sup>8,9</sup>, for building a recommendation for learners<sup>10</sup>, and for suggesting relevant discussions on a discussion board<sup>11</sup>. As the desire to use and understand learner behavior in OLEs increases, the need for analysis approaches not traditionally used for educational data becomes clear.

In this paper, we describe the application of a particular data analysis tool, association analysis, and demonstrate its relevance to data generated by learners in an OLE. Specifically, we demonstrate how association analysis can be used to: 1) better understand learner behavior on an OLE discovering common usage patterns in the site and 2) identify the learner interactions with the site that correspond to resultant learning of the site’s target learning objectives.

The first of these analyses generates rules of the type “users who visited page x and page y also visited page z” and shed light common usage patterns. They can be used both for better understanding how learners are interacting with an OLE and also for recommending particular pages. For examples, upon observing that a learner had visited page x and page y, in her active session window, the learner might be recommended to visit page z as well, similar to

recommendations shown on consumer sites of the type "consumers who viewed this item also bought the following".

The second application of association analysis that we present is based on discovering the relationships between visits to particular pages within the OLE and resultant learning. There are typically a large set of routes that can be taken in exploring a website, and because an OLE is usually designed with multiple learning objectives, it is valuable to discover the routes through the site that are associated with achieving each of these objectives. Of note is that the application of this analysis approach requires the measurement of user learning and so assumes that such data is available.

Consider an example of an OLE to learn basic arithmetic, where one goal is to teach the user the concept of fractions. The OLE may consist of several specific pages designated to teach fractions (say  $p_3, p_4, p_{74}, p_9$ ) as well as some other pages with pre-requisite material (say  $p_{65}, p_{98}, p_{101}$ ) and some examples ( $p_1, p_{60}$ ). Also, a question (say question 10) on an associated assessment measures the understanding of fractions. Then, the discovery of a rule such as "70% of users who visit a subset of the pages (say  $p_4, p_{65}, p_{101}, p_{60}$ ) score high on question 10 " is of value.

Of note is that approaches we describe can be used on any OLE that has available user tracking and associated user learning data. We demonstrate their application in a case study, which illustrates the methodology on the *CareerWISE* (CW) OLE, an online resource for teaching resilience skills to female STEM doctoral students. Specifically, we next discuss the general association analysis problem and then discuss the application of association analysis in an OLE including the data pre-processing and pattern discovery. We then describes the CW data set that serves as our example, and summarizes our findings and interpretations for this data set before drawing some conclusions and giving direction sfor future work.

### Association Analysis

Association analysis<sup>6,7</sup> is a popular data mining tool for discovering unordered relationships in large datasets. The general problem of association rule mining can be described as follows. Let  $I = \{ i_1, i_2, \dots, i_d \}$  denote a set of  $d$  binary attributes called items, a set of which is referred to as an itemset. Let  $T = \{ t_1, t_2, \dots, t_N \}$  denote the set of all  $N$  users' transactions each with a transaction ID where each  $t \in T$  is a subset of  $I$ . The transaction data can be presented with a binary representation as in Table 1. Here, an entry of one indicates that the item is present in the corresponding transaction, and a zero indicates it is absent.

**Table 1: A binary representation of transaction data for association analysis**

Transaction ID	$i_1$	$i_2$	...	$i_d$
1	1	0	...	1
2	1	1	...	0
...	0	0	...	1
$N$	1	0	...	0

The goal of the analysis is to discover relationships between the items in the form of association rules  $x \rightarrow y$ , where  $x$  and  $y$  are both itemsets. Here  $x$  is called the antecedent and  $y$  is the consequent of the rule. Such a rule implies that a transaction that contains the items in  $x$  tends to contain the item in  $y$ . Properties of association rules may be expressed in terms of different

defined measures. The *support count* of an itemset is the number of transactions that contain the itemset. We denote the number of elements in an itemset  $x$  by  $|x|$ . Then the support of itemset  $x$  is

$$\sigma(x) = |\{t_i \mid x \in t_i, t_i \in T\}|.$$

The *support* of a rule  $x \rightarrow y$ , is the proportion of transactions that contain the itemsets in the rule. This is the support count of the union of the antecedent and the consequent divided by the total number of transactions as in Equation 1.

$$s(x \rightarrow y) = \frac{\sigma(x \cup y)}{N} \quad (1)$$

Typically, one is interested in rules with large support (say, greater than 10%) because they occur in a substantial proportion of transactions. The *confidence* of a rule is the ratio of its support and the support of the antecedent as in Equation 2. The confidence is the proportion of transactions that contain the consequent among the transactions that contain the antecedent. One is interested in rules with large confidence (such as those greater than 80%).

$$c(x \rightarrow y) = \frac{\sigma(x \cup y)}{\sigma(x)} \quad (2)$$

It may also be useful to restrict our attention to rules that have specific items in their antecedent or consequent. Such a *syntactic constraint*<sup>7</sup> restricts the items that appear in the rule. For example, we may only be interested in a subset of rules where a specific item appears in the consequent (such as a question from an instrument).

A common strategy for mining association rules is to first find all itemsets that satisfy some pre-defined minimum support. In the next step, rules that satisfy a minimum confidence are generated from the frequent itemsets (any syntactic constraints are also considered in this process). The result of this approach is the generation of a subset of rules with syntactic constraints that satisfy some minimum support and confidence values. Generally, increasing the values for the minimum support, confidence and imposing syntactic constraints decreases the number of the generated rules.

Depending on the values of minimum support, confidence and the syntactic constraints, the described approach has the potential of generating a large number of rules. Hence, one usually relies on some criteria to filter for interesting rules<sup>12</sup>. It is important to note the limitation of assessing the strength of association rules solely based on the support and confidence measures: namely, a rule can have high confidence if the consequent of the rule has high support, even in the case of independence between the consequent and the antecedent.

As an example, suppose the evaluation of the rule *itemset 1*  $\rightarrow$  *itemset 2* is of interest. Table 2 depicts the support counts relating to these two itemsets<sup>12</sup>. The support and the confidence of this rule can be calculated to be 15% ( $=30/200$ , where 30 is the number of transactions in which both itemsets 1 and 2 were present, and 200 is the total number of transactions) and 75% ( $=30/40$ , where 40 is the total number of transactions that had itemset 1 present), making this rule seemingly interesting. However, a closer look reveals that the fraction of transactions that contain itemset 2, regardless of itemset 1, is 80% ( $=160/200$ , where 160 is the number of

transactions containing itemset 2). By comparing this value to the fraction of transactions that contain itemset 2 among the transactions that contain itemset 1 (75%), we observe that the presence of itemset 1 actually decreases the probability of the presence of itemset 2. This makes this rule uninteresting despite its high confidence. This necessitates the need for other measures for assessing the strength of an association rule. One such measure is the *lift*<sup>12</sup>.

**Table 2: An example of support counts relating to two itemsets. Data in the table shows the count for the number of transactions in which each of the itemsets was present or absent.**

Itemset 1	Itemset 2	
	Present in Transaction	Absent in Transaction
Present in Transaction	30	10
Absent in Transaction	130	30

The lift measure is the ratio of the rule's confidence and the support of the consequent, as shown in Equation 3.

$$l(x \rightarrow y) = \frac{c(x \rightarrow y)}{s(y)} \quad (3)$$

This measure is based on comparing the rule's confidence to a benchmark confidence based on the assumption that the occurrence of the consequent is independent of the occurrence of the antecedent. This benchmark confidence can be obtained from the support of the consequent. A lift value of 1 depicts independence between the antecedent and the consequent of the rule and a value greater than 1 depicts positive correlation. Generally the larger the lift ratio, the greater is the strength of the association.

In addition to the objective statistical criteria, the second set of criteria for filtering interesting rules is based on subjective arguments. These incorporate subjective knowledge from domain experts in judging the interestingness of the generated rules. Visualization can play an important role in this context, as it facilitates the domain expert in sifting through generated rules in search of interesting ones.

Using both objective and subjective criteria, we first generate rules that satisfy the minimum support and confidence value and syntactic constraints (if any), and then rely on the lift and a domain experts' knowledge for the final evaluation of interestingness of the rules which is facilitated by visualization tools.

### **Application of Association Analysis in an OLE**

The typical process for usage-based web analysis requires pre-processing the raw tracking data before the pattern discovery can begin<sup>13</sup>. Specifically, the data corresponding to learner's trajectories through the OLE must be transformed to a format similar to Table 1, in which each row of the data table represents a learner who visited the site, and each column corresponds to a particular page within the OLE. The binary data contained within the table corresponds to whether the learner visit the page. It is worth noting here that the tracking data retrieval and transformation process is greatly simplified if users are required to log-in with a unique identifier prior to accessing the OLE. After the transformation of the data to the required format, the

process of pattern discovery can begin using standard packages in freely available statistical software such as R.

Adopting the terminology described previously to our application, we let  $P = \{ p_1, p_2, \dots, p_d \}$  denote the set of binary attributes corresponding to the pages in the OLE. Each attribute  $p_j$  is referred to as a page and a set of these attributes is referred to as a *pageset*. The interaction of a user with a Website can be represented as a subset of  $P$ , where  $p_j = 1$  if the user has visited page  $j$  at least once. This will be referred to as the user's *navigation path*. We let  $NP = \{ np_1, np_2, \dots, np_N \}$  denote the set of all  $N$  users' navigation paths. Each user is assigned a user ID and has one navigation path. Then, an example of a relationship describing learner behavior on an OLE might be: "30% of users visit both page 1 and 5 and 80% of users who access page 1 also access page 5". More generally, rules in the form of  $\{ p_x, p_y, \dots \} \rightarrow p_z$ , are mined to discover the relationships between different pages.

The incorporation of syntactic constraints is also of interest in our application. For instance, it is of interest to mine rules that help explain the relationship between pagesets and the resultant user learning. This is presented in rules such as "70% of users who visit pageset  $x$  score high on question 10 in the learning instrument". Mining such consequent specific rules is achieved by augmenting the user's navigation path with new attributes corresponding to the question in the learning assessment and imposing the syntactic constraint to mine rules that have these attributes as their consequents. For instance, the score to each question can be binarized (two distinct values of high and low on a learning measure) and represented as a new binary attributes as in Table 3. In this table, a one in a column labeled as  $p_i$  indicates that the user visited that page, and the score of high or low on the question of interest is shown in the final two columns. By restricting the consequent of the rules to *score=high* to indicate that the user scored high on the question of interest, rules of the form  $\{ \text{visit to pageset} \} \rightarrow \text{score} = \text{high}$  can be mined. It should be noted that discretization of a numeric question is required to conform to the representation shown in Table 3. Although we have binarized the response into high and low, more than two categories can be used, and other algorithms to discretize can be applied<sup>14, 15</sup>.

**Table 3: A binary representation of OLE data relating the binarized score on a particular question in a learning instrument to pagesets within the OLE .**

User ID	$p_1$	$p_2$	...	$p_d$	<i>score=high</i>	<i>score=low</i>
1	1	0	...	0	1	0
2	0	0	...	1	0	1
...	...	...	...	...	...	...
$N$	1	0	...	1	1	0

## Case Study

To demonstrate the types of findings that can come of the application of association analysis to an OLE, we next present a case study illustrating the approach on the *CareerWISE* (CW) resilience training website (<http://careerwise.asu.edu>). The association analysis approach described previously was applied to the data generated by the CW site, and a summary of the results are presented.

## Data Description

The *CareerWISE* (CW) research program is a large, interdisciplinary research program housed at Arizona State University and supported by the National Science Foundation. One of the major components of the program is the development of an online, psycho-educational resilience training program designed to increase the persistence of women in Science, Technology, Engineering, and Math (STEM) doctoral programs<sup>16</sup>. The CW website consists of 238 unique pages, each classified in several ways including by problem-solving step (CW teaches a personal and interpersonal problem solving approach) and by problem type (difficulties with advisors, undesirable trade-offs between academic and personal responsibilities, unfriendly environments, and impediments to research and timely progress). The CW website users are able to interact with content that focuses on the application of a problem-solving model to manage personal and interpersonal difficulties that may be impeding their progress toward completing the PhD program.

A randomized controlled trial (RCT) of the first version of the CW intervention was performed to formally evaluate the effectiveness of the resource in affecting the variables it was designed to influence<sup>17</sup>. During the RCT, 133 women doctoral students in STEM programs from 27 universities around the United States were given access to the website and asked to spend at least five hours exploring the site in an unconstrained manner during a two-week time period. While participants explored the site, their behavior was tracked using a tracking mechanism with Drupal, which is the content management system on which the CW intervention is built. Following the exploration of the website, participants were given access to a final post-assessment, which measured key learning measures that CW is designed to influence. The assessment instrument includes different scales that reflect the social-psychological constructs such as Problem-Solving Knowledge, Resilience, and Coping Efficacy. Participants responded with levels of agreement to each of the items on the instrument (presented in a 5-point Likert scale from "Strongly Agree" to "Strongly Disagree"), and a few example items are provided below.

"I know how to identify the type of support I need from a mentor to help me meet my career goals."

"When things go wrong with my research I question my commitment to my field."

Of note is that further information on the methods, participants, and results of the *CareerWISE* RCT can be found elsewhere<sup>17</sup> and is outside the scope of this paper. We briefly describe the RCT study here simply to provide some context for the data upon which the case study's association analysis is to be applied.

## Results

The R software for statistical computing and graphics was used for the pattern discovery in the case study. The *apriori* algorithm<sup>6</sup> in R's *arules* package was implemented for the mining of association rules. The algorithm was used to generate all rules that met a minimum support and confidence with specific syntactic constraints (if any). The *arulesviz* packages was then used to summarize the rules graphically in an effort to facilitate the incorporation of subjective domain-expert knowledge in evaluating the interestingness of the rules.

To study the relationship between different pages within the CW site, rules were generated in the form of  $\{ p_x, p_y, \dots \} \rightarrow p_z$ . Specifically, all rules that had a minimum support of 30% and minimum confidence of 90% were generated, and the 50 rules with the highest lift are summarized graphically in Figure 1 and could be passed onto a domain expert for evaluation. This figure displays each rule as a circle with its incoming (antecedent) and outgoing (consequent) as arrows. The size and the grayscale of each circle depict the support and lift of the rule. Figure 1 shows that the 50 highest lift valued rules involve only 28 of the 238 pages of the OLE. It further depicts that  $p_{64}$ : assess-problem,  $p_{11}$ : climate,  $p_4$ : balance,  $p_2$ : advisor-issues, and  $p_{15}$ : delays, all at the center of the graphic, are shared as the antecedent of a large number of rules.

An example of such rules are given below:

$$\{ p_4: \textit{balance}, p_{15}: \textit{delays}, p_{37}: \textit{family friendly policies} \} \rightarrow p_{11}: \textit{climate}$$

This rule had a support of 0.3, meaning that  $p_4$ ,  $p_{15}$ ,  $p_{37}$  and  $p_{11}$  appeared together in 30% of the navigation paths. A confidence of 0.9 implies that when  $p_4$ ,  $p_{15}$ ,  $p_{37}$  were visited, 90% of time  $p_{11}$  was also visited. The lift of this rule is 1.7 since 53% of the navigation paths contain  $p_{11}$ . This implies that the probability of observing  $p_{11}$  in a user's navigation path is 1.7 times greater given that  $p_4$ ,  $p_{15}$ ,  $p_{37}$  and  $p_{11}$  are in her navigation path. The high confidence and the greater than one lift value make this rule interesting with regard to the objective interestingness measures. This rule conveys that users who visited pages (in any order) on balance, delays, and family friendly policies also visited pages related to work climate and is subjectively interesting as it reflects the need/interest of users to learn about balancing a career in science and a personal and family life in the context of a climate that is not always friendly to women.

As another example, consider the following rule:

$$\{ p_6: \textit{brainstorming}, p_{44}: \textit{how you think}, p_{74}: \textit{resources for graduate students} \} \rightarrow p_{68}: \textit{stereotype-threat}$$

This rule had a support of 0.3, confidence of 0.9, and lift of 1.8. The interpretation of these measures is similar to the first rule. The high confidence and lift value make this rule objectively interesting. This rule shows users who visited pages related to brainstorming, thinking, and resources for graduate student also visited the page related to stereo-type threat.

Our other interest is to demonstrate the feasibility of studying the relationship between pagesets of the OLE and resultant learning. To do so, rules in the form of  $\{ \textit{pageset} \} \rightarrow \textit{score} = \textit{high}$  are mined, which can be used in understanding the relationship between the different pages of the OLE and the learning outcomes. As discussed previously, an instrument that measures the learning is necessary for such an approach. In this case study, the CW post-assessment instrument is used. Rules can be mined to associate a scale score or a score on a particular question to user's navigation paths within the site. For illustration purposes, we present here some of the rules mined to relate user behavior on the site with score on this particular question in the post-assessment instrument: "When things go wrong with my research, I question my commitment to my field". The response to this question was binarized as high and low, and rules with the consequent of high score on this question were mined, indicating that the participants are not questioning their commitment to their field.

The apriori algorithm was used to generate all rules with a minimum support of 0.2 and minimum confidence of 0.7 that meet the syntactic constraint of the consequent being *score = high*. Figure 2 summarizes the top 50 highest lift valued rules and shows that the 50 highest lift valued rules for the question of interest involve only 16 of the 238 pages of the OLE. It further depicts that  $p_{44}$ : how you think,  $p_{65}$ : specify-concrete-problem,  $p_{69}$ : careerwise-problem-solving method review,  $p_2$ : advisor-issues,  $p_{66}$ : strategizing,  $p_{57}$ : setting-priorities are shared as the antecedent of a large number of rules.

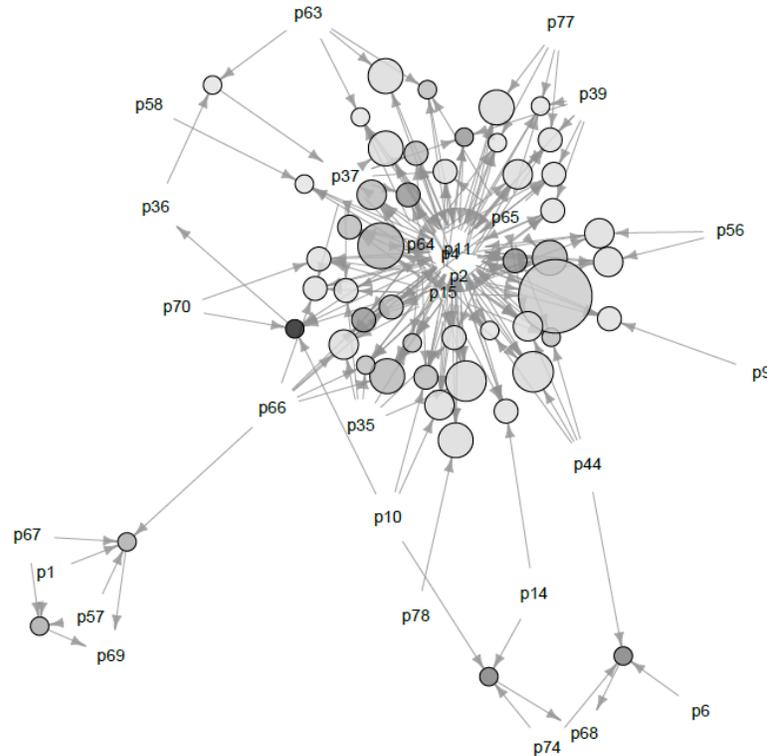


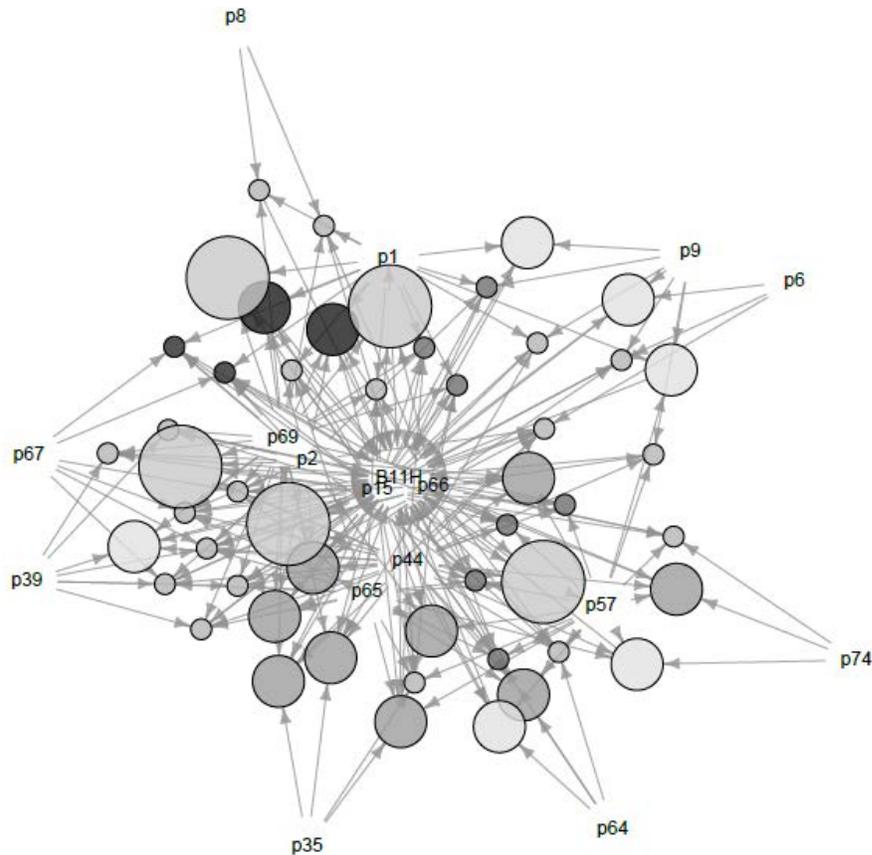
Figure 1: Graph-based representation of the page association rules: Each circle represents a rule. The incoming arrows of the circle depict the rule's antecedent and the outgoing arrow, the rule's consequent. The size and the grayscale of each circle depict the support (0.3-0.4) and lift (1.7-1.9) of the rule and  $p_j$  denotes page  $j$  of the CW OLE.

An example of a mined rule is given below:

$$\{ p_2: \text{advisor issues}, p_9: \text{career goals}, p_{15}: \text{delays} \} \rightarrow$$

*high score on "When things go wrong with my research I question my commitment to my field"*

This rule has a support of 0.2, confidence of 0.75 and a lift of 1.5. This rule demonstrates that participants who visited pages related to advisor issues, career goals, and delays, responded to the post-assessment saying they were not questioning their commitment to their field. As with the previous examples, this rule is both objectively and subjectively interesting. This information can be used to direct future site visitors who need to learn about this topic to the same pageset.



**Figure 2: Graph-based representation of the consequent specific associations rules:** Each circle represents a rule. The incoming arrows of a circle depict the rule's antecedent and the outgoing arrow, the rule's consequent (high score on the question of interest). The size and the grayscale of each circle depict the support (0.2-0.3) and lift (1.5-1.6) of the rule and  $p_j$  denotes page  $j$  of the CW OLE.

## Conclusion

This paper presented the application of two different association-based analyses that can be used for the understanding of how users interact with and learn from an existing OLE. We demonstrated how to mine rules that shed light on the users' learning interests and needs and also to depict the relationship between different pages of the OLE and the resulting learning outcome. These rules provide guidance on navigating the OLE and allow users to reach pages of the OLE that may have not been reached otherwise, which improves the OLE's efficiency at achieving targeted learning objectives. A case study applying the approaches to the *CareerWISE* OLE was given as an example, but it is important to note that the same approaches could be applied to any OLE for which the required data is available. Worth noting is that the work described here presents only a preliminary step in analyzing user interactions with an OLE. Further analyses should consider both the sequence of page visits as well as the length of time on each page.

Finally, of note is that this paper focused on presenting and demonstrating the feasibility of the association analysis methodology itself and not on the follow-up implications. Once the rules had been mined, they need to be presented to the educators and stakeholders associated with the OLE, who would be charged with determining whether and what corresponding changes to make to the OLE. For example, if a particular learning objective is considered very important, and the association analysis identifies a particular pageset as being associated with high learner scores on that topic, that identified page set could be explicitly recommended to learners prior to or upon first entry to the OLE. Future work in this area could include investigating the impact of such changes to an OLE such as the *CareerWISE* site.

## Acknowledgments

This study was supported by the National Science Foundation (NSF) grant 0910384. Any opinions, findings, and conclusions and recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the NSF.

## Bibliography

1. T.M. Harrison and T. Stephen. *Computer networking and scholarly communication in the twenty first century university*. State University of New York Press, 1996.
2. R. Kosala and H. Blockeel. Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, 2(1):1-15, 2000.
3. J. Srivastava, R. Cooley, M. Deshpande, and P.N. Tan. Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, 1(2):12-23, 2000.
4. Y. Zhang and J.R. Jiao. An associative classification-based recommendation system for personalization in b2c e-commerce applications. *Expert Systems with Applications*, 33(2):357-367, 2007.
5. R. Baker and K. Yacef. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1):3-17, 2009.
6. R. Agrawal, R. Srikant, et al. Fast algorithms for mining association rules. *In Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, volume 1215, pages 487-499, 1994.
7. R. Agrawal, T. Imieli\_nski, and A. Swami. Mining association rules between sets of items in large databases. *In ACM SIGMOD Record*, volume 22, pages 207-216. ACM, 1993.
8. O.R. Zaiane and J. Luo. Towards evaluating learners' behaviour in a web-based distance learning environment. *In Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference* , pages 357-360. IEEE, 2001.
9. R. Cooley, B. Mobasher, J. Srivastava, Data preparation for mining world wide web browsing patterns. *Knowledge and information systems*, 1(1):5-32, 1999.
10. O.R. Zaiane. Building a recommender agent for e-learning systems. In *Computers in Education, 2002. Proceedings. International Conference on*, pages 55-59. IEEE, 2002.
11. F. Abel, I. Bittencourt, N. Henze, D. Krause, and J. Vassileva. A rule-based recommender system for online discussion forums. *In Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 12-21. Springer, 2008.
12. P.N. Tan, M. Steinbach, V. Kumar. *Introduction to data mining*. Pearson Addison Wesley Boston, 2006.
13. B. Mobasher, R. Cooley, and J. Srivastava. Automatic personalization based on web usage mining. *Communications of the ACM*, 2000.
14. J. Dougherty, R. Kohavi, and M. Sahami. Supervised and unsupervised discretization of continuous features. *In Machine learning international workshop*, pages 194-202. Morgan Kaufmann Publishers, Inc., 1995.
15. U. Fayyad and K. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. 1993.

16. B. Bernstein. Managing barriers and building supports in science and engineering doctoral programs: Conceptual underpinnings for a new online training program for women. *Journal of Women and Minorities in Science and Engineering*, 17(1):29-50, 2011.
17. J.M Bekki, M.L. Smith, B. Bernstein, and C. Harrison. Effects of an online personal resilience training program for women in stem doctoral programs. *under review with the Journal of Women and Minorities in Science and Engineering*.