A Differential Temporal Interestingness Measure for Identifying the Learning Behavior Effects of Scaffolding

John S. Kinnebrew, Daniel L.C. Mack, and Gautam Biswas

Department of EECS and ISIS, Vanderbilt University, Nashville TN 37212, USA,
john.s.kinnebrew@vanderbilt.edu

Abstract. Effective design and improvement of scaffolding in complex and open-ended learning environments, requires the ability to assess the effectiveness of a variety of scaffolding options, not only in terms of overall performance and learning, but also in terms of more subtle effects on students' behavior and understanding. In this paper, we present a novel data mining technique that aids the analysis of scaffolding and students' learning behaviors by identifying activity patterns that distinguish groups of students (e.g., groups that received different scaffolding and feedback during an extended, complex learning activities) by differences in both total behavior pattern usage and evolution of pattern usage over time. We demonstrate the utility of this technique through application to student activity data from a recent experiment with the Betty's Brain learning environment and four different scaffolding conditions.

Keywords: learning behaviors, interestingness measure, sequence mining, information gain

1 Introduction

In order to more effectively teach and promote skills required in the modern world of near-ubiquitous computing and internet connectivity, computer-based learning environments have become more complex and open-ended. This complexity also drives a need for dynamic and adaptive scaffolding that can support students in understanding how to employ and learn with these environments and tools. However, in order to effectively design and improve such scaffolding, we must first be able to assess the effectiveness of a variety of scaffolding options, not only in terms of overall performance and learning, but also in terms of more subtle effects on students' behavior and understanding. In this paper, we present a novel data mining technique that aids the analysis of how students' learning behaviors and strategies are employed with differing frequency over the course of learning or problem-solving activities as the result of different scaffolds and feedback that can be provided in a learning environment.

Identifying sequential patterns in learning activity data can be useful for discovering and understanding student learning behaviors. Researchers have applied sequence mining techniques to a variety of educational data in order to better understand learning. The primary sequential pattern mining task is to discover sequential patterns of items that are found in many of the sequences in a given dataset [1, 2]. For example, Perera *et al.* ([3]) use sequential pattern mining to provide mirroring and feedback tools to support effective teamwork among students collaborating on software development using an open source professional development environment called TRAC. Other researchers have also employed sequential pattern mining to identify differences among student groups or generate student models to customize learning to individual students [4–6].

Once these behavior patterns are mined, researchers must interpret and analyze the resulting patterns to identify a relevant subset of important patterns that provide a basis for generating actionable insights about how students learn, solve problems, and interact with the environment. Researchers have developed a variety of measures to utilize properties other than the default of pattern frequency to rank mined patterns [7]. These measures are often referred to as "interestingness measures" and have been applied data mining tasks like sequence mining and association rule mining [8]. To better analyze student learning and behavior, interestingness measures have been used for tasks such as ranking mined association rules (e.g., [9]).

Investigation of the frequency with which a pattern occurs over time can reveal additional information for pattern interpretation and may help identify more important patterns, which occur only at certain times or become more/less frequent, rather than patterns with frequent, but uniform, occurrence over time. In this paper, we present a novel approach, combining sequence mining and an information-theoretic measure for ranking behavior patterns that distinguish groups of sequences (e.g., groups of students in different experimental conditions) by differences in both total pattern usage and the evolution of pattern usage over time. To effectively analyze these patterns and quickly identify trends in the evolution of pattern usage, we employ a related visualization in the form of heat maps.

2 Identifying Interesting Differences in Pattern Usage

In this section, we present the Differential Temporal Interestingness of Patterns in Sequences (D-TIPS) technique, and its novel interestingness measure, for identifying and visualizing patterns that are employed differentially over time among groups of students (e.g., groups that receive different scaffolding in an open-ended learning environment). The first step in analyzing learning activity sequences is to define and extract the actions that make up those sequences from interaction traces logged by the environment. The definition of actions in these sequences for Betty's Brain data is discussed further in Section 3. Given a set of sequences corresponding to the series of actions performed by each student, the D-TIPS technique consists of four primary steps:

1. Generate candidate patterns that are common to the majority of students in at least one group by combining the sets of patterns identified through ap-

plying sequential pattern mining separately to each group's learning activity sequences (with a frequency threshold of 50%).

2. Calculate a temporal footprint for each candidate pattern by mapping it back to locations where it occurs in the activity sequences. Specifically, each sequence is divided into n consecutive slices, such that each contains $\frac{100}{n}\%$ of the student's actions in the full sequence, where n is the chosen number of bins defining the temporal granularity of the comparisons. Corresponding slices for a group (e.g., the first slice from each sequence in the group, the second slice from each, and so on) are then grouped into bins and each action in the slices is marked to indicate whether or not it is the beginning of a pattern match in its original sequence. This set of binned and marked actions defines the temporal footprint of the pattern for the group.
3. Provide a ranking of the candidate patterns using an information-theoretic interestingness measure (described in more detail below) applied to the temporal footprint of each pattern.
4. For the highly-ranked patterns, visualize their temporal footprints using heat maps to identify differences in usage trends and spikes across student groups. Specifically, we employ a two-dimensional heat map where the y-axis is student group and the x-axis is time discretized by temporal bin. In a single row (i.e., for a specific student group), each cell's count is the percentage of total pattern occurrence (with respect to the student group) within the corresponding temporal bin. The use of *percentages* of pattern occurrence allows analysis of temporal variation normalized by the total frequency of the pattern in the group, which will tend to highlight different temporal trends in pattern usage across groups, even when total pattern occurrence differs significantly among groups.

In order to identify more interesting patterns by their difference in temporal usage across groups in step 3, the D-TIPS interestingness measure applies information gain (IG) with respect to pattern occurrence across the groups in each of the n corresponding bins of their temporal footprints. Information gain is defined as the difference in expected information entropy [10] between one state and another state where some additional information is known (e.g., the difference between a set of data points considered as a homogeneous group versus one split into multiple groups based on the value of some other feature or attribute). Information entropy is the amount of expected uncertainty found in a random variable, X, whose value can be called the *class* of the data point. IG when used in classifiers, such as decision trees, is applied to a dataset where each data point has multiple features in addition to its class. The IG of a given feature is then the reduction in expected uncertainty about the correct class of a data point when its feature value is known, or conversely the increase in information about the class of the data point. IG is calculated as the difference between the information entropy of the data without knowledge of the feature values and the conditional information entropy when the feature values are known.

Information gain is leveraged in classifiers to determine which features are most discriminatory because they provide the least amount of uncertainty among classes in the data. D-TIPS applies information gain to determine which patterns are the most interesting because knowledge of their occurrence and temporal location provides the least amount of uncertainty among the student groups. In D-TIPS, each action/data-point's class is its group, and the feature of each data point, for a given pattern, is the combination of whether the action begins an occurrence of the pattern *and* the number of the bin in which the action occurred. This information-theoretic definition of the D-TIPS measure provides two important properties: 1) given two patterns with the same total occurrences for corresponding groups, the pattern with the greater discrimination of groups by *differences in temporal location/bin among groups* will have a higher rank, and 2) given two patterns with the same relative temporal behaviors (i.e., the same proportion of total pattern occurrence in each bin) for corresponding groups, the pattern with the greater discrimination of groups by *differences in total occurrence among groups* will have a higher rank.

Therefore, the D-TIPS measure provides a way of recognizing differences among groups both by total pattern occurrence and by temporal behavior (e.g., decreasing usage versus increasing usage, or spikes in different bins). Further, when the same differences across groups occur for two patterns, the pattern with higher overall frequency will have the higher rank. Thus, D-TIPS tends to emphasize patterns with large relative differences among groups (by total occurrence and/or temporal behavior) even when they are not especially frequent in the overall dataset, while also emphasizing patterns with more moderate differences among groups when the frequency of the pattern in the overall dataset is high. Conversely, D-TIPS tends to deemphasize patterns that are homogeneous across groups (by both relative occurrence and temporal behavior) or that are especially rare in all groups.

3 Betty's Brain Data

The data we employ in the analysis in Section 4 consists of student interaction trace from the Betty's Brain [11] learning environment. In Betty's Brain, students read about a science process and teach a virtual agent about it by building a causal map. They are supported in this process by a mentor agent, who provides feedback and support for their learning activities. The data analyzed here was obtained in a recent study with 68 7^{th}-grade students taught by the same teacher in a middle Tennessee school. At the beginning of the study, students were introduced to the science topic (global climate change) during regular classroom instruction, provided an overview of causal relations and concept maps, and given hands-on training with the system. For the next four 60-minute class periods, students taught their agent about climate change and received feedback on content and learning strategies from the mentor agent.

The study tested the effectiveness of two support modules designed to scaffold students' understanding of cognitive and metacognitive processes important for success in Betty's Brain (details provided in [12]). The *knowledge construction* (KC) support module scaffolded students' understanding and suggested strategies on how to construct knowledge by identifying causal relations in the resources, and the *monitoring* (Mon) support module scaffolded students understanding and suggested strategies on how to monitor Betty's progress by using the quiz results to identify correct and incorrect causal links on Betty's map. Participants were divided into a control and three treatment groups. The knowledge construction (KC) group used a version of Betty's Brain that included the KC support module and a causal link tutorial that they could access at any time and were prompted to enter when the mentor determined they were having difficulty identifying causal links in the resources. The monitoring (Mon) group used a version of Betty's Brain that included the Mon support module and a tutorial about employing link annotations to keep track of links shown to be correct by quizzes. The full (Full) group used a version of Betty's Brain that included both support modules and tutorials. Finally, the control (Con) group used a version that included neither the tutorials nor the support modules.

In Betty's Brain, the students' learning and teaching tasks were organized around seven activities: (1) reading resource pages to gain information, (2) adding or removing causal links in the map to organize and teach causal information to Betty, (3) querying Betty to determine her understanding of the domain based on the causal map, (4) having Betty take quizzes that are generated and graded by the mentor to assess her current understanding and the correctness of links in the map, (5) asking Betty for explanations of which links she used to answer questions on the quiz or queries, (6) taking notes for later reference, and (7) annotating links to keep track of their correctness determined by quizzes and reading. Actions were further distinguished by context details, which for this analysis were the correctness of a link being edited and whether an action involved the same subtopic of the domain as at least one of the previous two actions. The definition of actions in Betty's Brain learning activity sequences are discussed further in [13].

4 Results

To illustrate and characterize the performance of the D-TIPS technique on educational data, we present selected results from its application to student learning activity data in the Betty's Brain classroom study described in Section 3. The D-TIPS analysis identified 560 activity patterns that occurred in at least half of the students in one or more of the four experimental conditions. Given the limited number of students in each condition, we chose to bin pattern occurrence values into fifths of the activity sequences for a broad analysis of their usage evolution over time. Table 1 presents 3 of the top 30 most differentially-interesting patterns identified by D-TIPS across the four scaffolding conditions. For comparison, the average occurrences per student and ranking by that value is also presented. Over half (18) of the 30 analyzed D-TIPS patterns had a rank past 50th by occurrence, with 13 of them ranking beyond 100th, indicating that they would be unlikely to have been considered without D-TIPS.

Table 1. Selected Patterns with D-TIPS and Occurrence Rankings

Pattern	D-TIPS Rank	Occurrence Rank	Avg Occurrence
[Quiz]	3	2	21.8
[Read] → [Note]	18	100	1.7
[Read] → [Read] → [Remove Link$^-$]	29	137	1.4

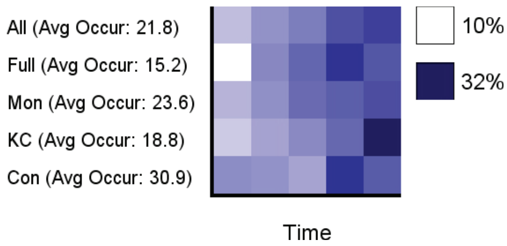

Fig. 1. [Quiz]

The first pattern in Table 1 illustrates a single action pattern that was ranked very high by both D-TIPS and overall occurrence. While individual student actions are often less interesting than longer patterns, they are still important to consider, especially when they also illustrate a tendency to be employed differentially across groups and over time. Figure 1 shows that all groups tended to use quizzes more frequently later in their work on the system. Since students' causal maps grew over time, monitoring and correction of the maps were more important later in their learning activities. There were some differences in usage trends over time among the different conditions, such as the steeper increasing trend for the KC and Full groups than the Monitoring group and the earlier peak in usage for the Full and Control groups. However, the overall occurrence by conditions differed markedly, with the Control group performing far more quiz actions than the others, and the Monitoring group performing more quiz actions than the KC and Full groups. While the Monitoring group's use of the quiz was expected to be high due to the focused monitoring support that relied heavily on the quiz, it is surprising that the Control group had the highest quiz usage. This might indicate that without either KC or monitoring support, the Control group struggled more and fell back on guessing and checking (with the quiz) strategies.

Figure 2 illustrates a knowledge construction behavior of reading and taking notes that was ranked highly by D-TIPS. Another difference among the groups, which added to the interestingness of this pattern under the D-TIPS analysis, is that the Control group tended to perform reading followed by note-taking primarily in the last fifth of their activities, as opposed to the first two fifths for the other groups. However, further analysis of the data attributed this primarily to only two of the Control group students, although the reason for this aberration is still unclear.

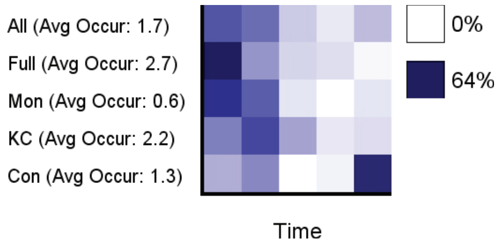

Fig. 2. [Read] → [Note]

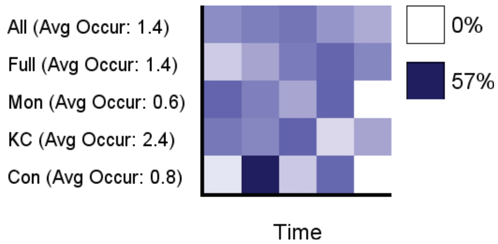

Fig. 3. [Read] → [Read] → [Remove Link$^-$]

The pattern illustrated in Figure 3 involves a sequence of (two) reading actions followed by removing an incorrect link. While there was no consistent temporal trend in the usage of this pattern, the Monitoring and Control groups exhibited this pattern less than once per student, while the KC group averaged 2.4 times per student. Although ranked lower by D-TIPS at 45th, the sub-pattern of a single read action followed by removing an incorrect link illustrates the same differences. This suggests that students with the KC feedback, relied more heavily on reading to identify incorrect links than either the Control and Monitoring groups, possibly because the Control group struggled more in general and the support in the Monitoring group focused students more on the use of quizzes to identify incorrect links.

5 Conclusion

While identification of high-frequency patterns is undoubtedly useful, finding patterns that have differing usage over time across a set of student groups is also important for analyzing the effects of scaffolding. In this paper, we presented the D-TIPS technique, which identifies patterns that differ in their usage among student groups by either total (group) occurrence or temporal behavior, even when they are not especially frequent in the overall dataset. Results from the use of this technique to mine Betty's Brain data illustrated the potential benefits and helped characterize differences between D-TIPS and a baseline occurrence ranking. D-TIPS identified patterns that illustrated potentially important differences in learning behavior among different scaffolding conditions that would have probably been overlooked by considering only pattern frequency. Future work will include autonomous identification of an effective number of bins for splitting a given set of activity sequences, as well as methods to individually characterize student groups by the patterns identified in D-TIPS.

6 Acknowledgments

This work has been supported by NSF-IIS Award #0904387 and IES Award #R305A120186.

References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceedings of the Eleventh IEEE International Conference on Data Engineering (ICDE). (1995) 3–14
2. Zaki, M.: Sequence mining in categorical domains: incorporating constraints. In: Proceedings of the ninth international conference on Information and knowledge management, ACM (2000) 422–429
3. Perera, D., Kay, J., Koprinska, I., Yacef, K., Zaïane, O.: Clustering and sequential pattern mining of online collaborative learning data. IEEE Transactions on Knowledge and Data Engineering **21**(6) (2009) 759–772
4. Amershi, S., Conati, C.: Combining unsupervised and supervised classification to build user models for exploratory learning environments. Journal of Educational Data Mining **1**(1) (2009) 18–71
5. Martinez, R., Yacef, K., Kay, J., Al-Qaraghuli, A., Kharrufa, A.: Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In: Proceedings of the Fourth International Conference on Educational Data Mining, Eindhoven, Netherlands (2011)
6. Su, J.M., Tseng, S.S., Wang, W., Weng, J.F., Yang, J., Tsai, W.N.: Learning portfolio analysis and mining for scorm compliant environment. Journal of Educational Technology and Society **9**(1) (2006) 262–275
7. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: A survey. ACM Computing Surveys (CSUR) **38**(3) (2006) 9
8. Zhang, Y., Zhang, L., Nie, G., Shi, Y.: A survey of interestingness measures for association rules. In: Business Intelligence and Financial Engineering, 2009. BIFE'09. International Conference on, IEEE (2009) 460–463
9. Merceron, A., Yacef, K.: Interestingness measures for association rules in educational data. Educational Data Mining 2008 (2008) 57
10. Renyi, A.: On measures of entropy and information. In: Fourth Berkeley Symposium on Mathematical Statistics and Probability. (1961) 547–561
11. Biswas, G., Leelawong, K., Schwartz, D., Vye, N., Vanderbilt, T.: Learning by teaching: A new agent paradigm for educational software. Applied Artificial Intelligence **19**(3) (2005) 363–392
12. Biswas, G., Kinnebrew, J.S., Segedy, J.R.: Analyzing students' metacognitive strategies in open-ended learning environments. In: Proceedings of the 35th annual meeting of the Cognitive Science Society, Berlin, Germany (August 2013)
13. Kinnebrew, J.S., Loretz, K.M., Biswas, G.: A contextualized, differential sequence mining method to derive students' learning behavior patterns. Journal of Educational Data Mining (In Press, 2013)