Towards Using Coherence Analysis to Scaffold Students in Open-Ended Learning Environments

James. R. Segedy & Gautam Biswas

Institute of Software Integrated Systems, Department of Electrical Engineering and Computer Science, Vanderbilt University, 1025 16th Avenue South, Nashville, TN, 37212, U.S.A.
{james.segedy,john.s.kinnebrew,gautam.biswas}@vanderbilt.edu

Abstract. Scaffolding students in open-ended learning environments (OELEs) is a difficult challenge. The open-ended nature of OELEs allows students to simultaneously pursue, modify, and abandon any of a large number of both short-term and long-term approaches to completing their tasks. To overcome these challenges, we have recently developed *coherence analysis*, which focuses on students' ability to *interpret* and *apply* the information available in the OELE. This approach has yielded valuable dividends: by characterizing students according to the coherence of their behavior, teachers and researchers have access to easily-calculated, intuitive, and actionable measures of the *quality* of students' problem-solving processes. The next step in this line of research is to develop a framework for using coherence analysis to adaptively scaffold students in OELEs. In this paper, we present our initial ideas for this work and propose guidelines for the construction of a scaffolding framework.

Keywords: Open-ended learning environments, metacognition, coherence analysis, scaffold

1 Introduction

Open-ended computer-based learning environments (OELEs) [1-2] are learner-centered; they present students with a challenging problem-solving task, information resources, and tools for completing the task. Students must use the resources and tools to construct and verify problem solutions, and in this process learn about the problem domain and develop their general problem-solving abilities. In OELEs, students have to distribute their time and effort between exploring and organizing their knowledge, creating and testing hypotheses, and using their learned knowledge to create solutions. Since there are no prescribed solution steps, students may have to discover the solution process over several hours. For example, learners may be given the following:

> *Use the provided simulation software to investigate which properties relate to the distance that a ball will travel when rolled down a ramp, and then use what you learn to design a wheelchair ramp for a community center.*

Whereas OELEs support a constructivist approach to learning, they also place significant cognitive demands on learners. To solve problems, students must simultaneously wrestle with their emerging understanding of complex topics, develop and utilize skills to support their learning, and employ *self-regulated learning* (SRL) processes to manage the open-ended nature of the task. SRL is a theory of learning that describes how learners actively set goals, create plans for achieving those goals, continually monitor their progress, and revise their plans when necessary to continue to make progress [3]. As such, OELEs can *prepare students for future learning* [4] by developing their ability to independently investigate and develop solutions for complex open-ended problems.

However, research with OELEs has produced mixed results. While some students with higher levels of prior knowledge and SRL skills show large learning gains as a result of using OELEs, many of their less capable counterparts experience significant confusion and frustration [5-7]. Research examining the activity patterns of those students indicates that they typically make ineffective, suboptimal learning choices when they independently work toward completing open-ended tasks [7-10].

The strong self-regulatory component of OELEs makes them an ideal environment for studying SRL. The open-ended nature of the environment forces students to make choices about how to proceed, and these choices reveal information about students' understanding of: (i) the problem domain; (ii) the problem-solving task; and (iii) strategies for solving the problem. By studying these choices, we can gain a better understanding of how students regulate their learning and how best to design scaffolds to support students who struggle to succeed.

Recently, we have introduced *coherence analysis* (CA) [11], a technique for studying students' problem-solving behaviors in OELEs. CA analyzes learners' behaviors in terms of their demonstrated ability to seek out, interpret, and apply information encountered while working in the OELE. By characterizing behaviors in this manner, CA provides insight into students' problem-solving strategies as well as the extent to which they understand the nuances of the learning and problem solving tasks they are currently completing.

In this paper, we present an overview of our findings with coherence analysis as applied to the *Betty's Brain* OELE (REF) and present our plans on extending this research. Our goal with CA is to empower both human and virtual tutors to more powerfully support students as they learn complex open-ended problem solving.

2 Betty's Brain

Betty's Brain [11] presents the task of teaching a virtual agent, Betty, about a science phenomenon (e.g., climate change) by constructing a causal map that represents that phenomenon as a set of entities connected by directed links representing causal relationships. Once taught, Betty can use the map to answer causal questions. The goal for students is to construct a causal map that matches an expert model of the domain.

In *Betty's Brain*, students acquire domain knowledge by reading resources that include descriptions of scientific processes (*e.g.*, shivering) and information pertaining to each concept that appears in the expert map (*e.g.*, friction). As students read, they need to identify causal relations such as "*skeletal muscle contractions create friction in the body.*" Students can then apply this information by adding the entities to the map and creating a causal link between them (which "teaches" the information to Betty). Learners are provided with the list of concepts, and link definitions may be either increase (+) or decrease (-).

Learners can assess their causal map by asking Betty to answer questions and explain her answers. To answer questions, Betty applies qualitative reasoning to the causal map (*e.g.*, *the question said that the hypothalamus response increases. This causes skin contraction to increase. The increase in skin contraction causes…*). After Betty answers a question, learners can ask Mr. Davis, another pedagogical agent that serves as the student's mentor, to evaluate her answer. If Betty's answer and explanation match the expert model (*i.e.*, in answering the question, both maps utilize the same causal links), then Betty's answer is correct.

Learners can also have Betty take *quizzes* (by answering sets of questions). Quiz questions are selected dynamically by comparing Betty's current causal map to the expert map such that a portion of the chosen questions, in proportion to the completeness of the current map, will be answered correctly by Betty. The rest of her quiz answers will be incorrect or incomplete, helping the student identify areas for correction or further exploration. When Betty answers a question correctly, students know that the links she used to answer that question are correct. Otherwise, they know that at least one of the links she used to answer the question is incorrect. Students may keep track of correct links by annotating them as such.

3 Coherence Analysis

The Coherence Analysis (CA) approach analyzes learners' behaviors by combining information from sequences of student actions to produce measures of *action coherence*. CA interprets students' behaviors in terms of the information they encounter in the OELE and whether or not this information is utilized during subsequent actions. When students take actions that put them into contact with information that can help them improve their current solution, they have *generated potential* that should *motivate future actions*. The assumption is that if students can recognize relevant information in the resources and quiz results, then they should act on that information. If they do not act on information that they encountered previously, CA assumes that they did not recognize or understand the relevance of that information. This may stem from incomplete or incorrect understanding of the domain under study, the learning task, and/or strategies for completing the learning task. Additionally, when students add to or edit their problem solution when they have not encountered any information that could motivate that edit, CA assumes that they are guessing[1]. These two notions come together in the definition of action coherence:

[1] Students may be applying their prior knowledge, but the assumption is that they are novices to the domain and should verify their prior knowledge during learning.

Two ordered actions ($x \rightarrow y$) taken by a student in an OELE are ***action coherent*** *if the second action, y, is based on information generated by the first action, x. In this case, x provides* ***support*** *for y, and y is* ***supported*** *by x. Should a learner execute x without subsequently executing y, the learner has created* ***unused potential*** *in relation to y. Note that actions x and y need not be consecutive.*

CA assumes that learners with higher levels of action coherence possess stronger metacognitive knowledge and task understanding. Thus, these learners will perform a larger proportion of supported actions and take advantage of a larger proportion of the potential that their actions generate. In the analyses performed to date, we have incorporated the following coherence relations:

- Accessing a resource page that discusses two concepts *provides support for* adding, removing, or editing a causal link that connects those concepts.
- Viewing assessment information (usually quiz results) that proves that a specific causal link is correct *provides support for* adding that causal link to the map (if not present) and annotating it as being correct (if not annotated).
- Viewing assessment information (usually quiz results) that proves that a specific causal link is incorrect *provides support for* deleting it from the map (if present).

Using these coherence relations, we derived six primary measures describing students' problem solving processes:

1. *Edit Frequency*: The number of causal link edits and annotations made by the student per minute on the system.
2. *Unsupported edit percentage*: the percentage of causal link edits and annotations not supported by information encountered within 5 minutes of the edit/annotation.
3. *Information viewing time*: the amount of time spent viewing either the science resources or Betty's graded answers. *Information viewing percentage* is the percentage of the student's time on the system classified as *information viewing time*.
4. *Potential generation time*: the amount of *information viewing time* spent viewing information that could support causal map edits that would improve the map. To calculate this, we annotated each hypertext resource page with information about the concepts and links discussed on that page. *Potential generation percentage* is the percentage of *information viewing time* classified as *potential generation time*.
5. *Used potential time*: the amount of *potential generation time* associated with information viewing that both occurs within a prior five minute window of and also supports an ensuing causal map edit. *Used potential percentage* is the percentage of *potential generation time* classified as *used potential time*.
6. *Disengaged time*: the sum of all periods of time, at least five minutes long, during which the student neither viewed a source of information for at least 30 seconds nor edited the map. *Disengaged percentage* is the percentage of the student's time on the system classified as disengaged time.

Metrics one and two capture the quantity and quality of a student's causal link edits and annotations, where supported edits and annotations are considered to be of higher quality. Metrics three, four, and five capture the quantity and quality of the student's time viewing either the resources or Betty's graded answers. These metrics speak to the student's ability to seek and identify information that may help them build or refine their map (potential generation percentage) and then utilize information from those pages in future map editing activities (used potential percentage). Metric 6 represents periods of time during which the learner is not measurably engaged with the system.

3.1 Summary of Findings with Coherence Analysis

Coherence analysis has proved to be a valuable tool for understanding how students learn as they solve open-ended problems. Thus far, we have investigated it with one group of 98 6th-grade students (11 year olds). Thus, we interpret our findings with cautious optimism. We have identified the following relationships:

- *CA predicts learning and performance*: in general, students with higher levels of coherent behaviors have shown significantly higher levels of success in teaching Betty. Moreover, these learners have shown a better understanding of the science domain they were learning [11].
- *Prior skill levels predict CA*: students who were better able to identify causal links in abstract text passages (*e.g.*, A decrease in Ticks leads to an increase in Tacks) exhibited higher levels of coherence while using *Betty's Brain* [11].
- *CA identifies common problem solving profiles across students*: we clustered students by describing them with the six CA metrics described above, and we identified five common profiles among students: researchers and careful editors; strategic experimenters; confused guessers; disengaged students; engaged and efficient students. Interestingly, there were few differences in learning and performance among the clusters. Engaged and efficient students showed higher learning and performance than the other clusters, but there were not any other meaningful differences, suggesting that CA allows us to understand how different learning approaches lead to similar learning outcomes [11].
- *CA identifies common day-to-day problem solving profiles and transitions among them*: we clustered students as before, but this time the unit of analysis was a single day of using the system instead of the entire time using the system. We found a set of behavior profiles quite similar to those identified in the previous analysis. In analyzing day-to-day transitions, we found that many students performed fairly consistently while several other students performed inconsistently (that is, they have days of high coherence and days of low coherence). We also identified common transitions among days, which allowed us to find a potentially *at-risk* behavior profile. Students who behave like researchers and careful editors are far more likely than chance to transition to confused or disengaged behavior in subsequent days [12].

4 An Initial Coherence-Based Scaffolding Framework

Given the previous findings with CA, we aim to utilize the power of the analysis in real time as students use the system in order to detect non-coherent behavior, diagnose the cause of it, and take steps to support students in overcoming the difficulties they are experiencing. The core idea behind CA is that when students work in OELEs, they have two primary sets of tasks: *information seeking tasks* related to identifying and interpreting important information and *information application tasks* related to applying that information to improving the problem solution. All coherence metrics are based on identifying relationships between activities related to these two sets of tasks. By analyzing student behaviors with CA, we can identify problems related to information seeking and information application.

4.1 Diagnosing Problems with CA Metrics

The initial framework for diagnosing problems using CA metrics appears in Figure 1. This framework maps CA metrics to the problems they may indicate. For example, low levels of potential generation indicate that the learner is spending a large portion of their information viewing time on non-helpful information. This indicates that they may be struggling to identify relevant vs. non-relevant information in the environment. Problems with information seeking may also manifest as high levels of unused potential (*i.e.*, not applying viewed information), a high proportion of unsupported edits, and a low rate of editing the solution. Problems with information application are indicated by high unused potential and a low rate of editing the solution.

CA metrics may also be used to identify behaviors associated with effort avoidance. Specifically, low levels of information viewing, a low rate of editing the solution, a high unsupported edit percentage, and high levels of disengagement indicate that the learner may be purposefully avoiding effort. This may be due to a number of reasons, including low self-efficacy and low skill understandings.

Using this framework, our initial plan for using CA to scaffold students is as follows:

1. Observe the student for a period of time (*e.g.*, 10 minutes) and calculate their coherence metrics for that period. Identify any problematic behaviors (*e.g.*, high unused potential).
2. Form hypotheses about the sources of these behaviors. This involves looking at the combination of problematic behaviors observed and the student's previous activities in the system. For example, if the problematic behaviors are high unused potential and a low editing rate, the system may hypothesize that the student is struggling to apply information.

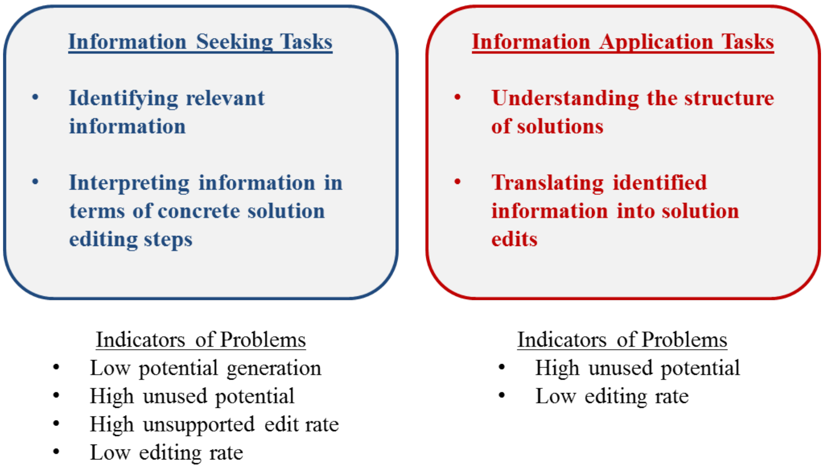

Fig. 1. Initial Problem Diagnosis Framework

3. Perform active diagnosis of the student to resolve competing hypotheses and gain additional information. For example, if the student has a high unsupported edit rate, this may be due to effort avoidance or a misunderstanding related to information seeking. The system can have the student answer questions and complete short problems in order to gain additional evidence as to which of these is the actual problem.
4. Once the system is confident that the student is struggling to understand something, it can use *guided practice scaffolds* [13] to help the student learn the knowledge and skills that they are missing or about which they are confused. Throughout guided practice, the system should provide encouragement, feedback, and scaffolding. It should also reinforce the relevance of the targeted knowledge and skills to the primary problem solving task, problem solving in general, and academic success.
5. If the system is confident that the student is exhibiting effort avoidance, then it should offer to help the student. If the behavior continues after the offer (and potential scaffolding related to that offer), then the system should provide guided practice scaffolds on the important knowledge and skills they need to understand to be successful. Hopefully, the student's abilities will improve during guided practice, and that will re-engage them with the learning task. As in the previous step, the system should provide the student with encouragement, feedback, and scaffolding and it should reinforce the relevance of the targeted knowledge and skills.

5 Conclusion

In this paper, we have provided an overview of *coherence analysis* (CA), an analysis approach that provides insight into how students behavior in open-ended computer-based learning environments (OELEs). Additionally, we have presented an initial scaffolding framework that describes how CA might be leveraged to provide adaptive scaffolds to students who are struggling. As we move forward, we will continue developing this scaffolding framework, build it into *Betty's Brain*, and test its effectiveness with students.

Acknowledgements. This work has been supported by Institute of Education Sciences CASL Grant #R305A120186.

References

1. Land, S., Hannafin, M., Oliver, K.: Student-centered learning environments: Foundations, assumptions and design. In: D. Jonassen & S. Land (eds.) Theoretical Foundations of Learning Environments, pp. 3-25. Routledge, New York, NY (2012)
2. Segedy, J.R., Biswas, G., Sulcer, B.: A model-based behavior analysis approach for open-ended environments. The Journal of Educational Technology & Society, 17(1), 272-282 (2014)
3. Zimmerman, B., Schunk, D. (eds.): Handbook of Self-Regulation of Learning and Performance. Routledge , New York, NY (2011)
4. Bransford, J., Schwartz, D.: Rethinking transfer: A simple proposal with multiple implications. Review of Research in Education, 24(1), 61-101 (1999)
5. Azevedo, R., Witherspoon, A.: Self-regulated learning with hypermedia. In: Hacker, D., Dunloskey, J., Graesser, A. (eds.), Handbook of Metacognition in Education, pp. 319-339. Taylor and Francis (2009)
6. Hacker, D., Dunloskey, J., Graesser, A. (eds.), Handbook of Metacognition in Education, pp. 319-339. Taylor and Francis (2009)
7. Kinnebrew, J.S., Loretz, K., Biswas, G.: A contextualized, differential sequence mining method to derive students' learning behavior patterns. Journal of Educational Data Mining, 5(1), 190-219 (2013)
8. Land, S.M.: Cognitive requirements for learning with open-ended learning environments. Educational Technology Research and Development*,* 48, 61-78 (2000)
9. Mayer, R.E.: Should there be a three-strikes rule against pure discovery learning?, American Psychologist, 59, 14-19 (2004)
10. Sabourin, J., Shores, L., Mott, B., Lester, J.: Understanding and predicting student self-regulated learning strategies in game-based environments. International Journal of Artificial Intelligence in Education, 23, 94-114 (2013)
11. Segedy, J.R., Kinnebrew, J.S., Biswas, G.: Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. Journal of Learning Analytics (in press)
12. Segedy, J.R., Biswas, G.: Coherence over time: Understanding day-to-day changes in students' open-ended problem solving behaviors. In: Proceedings of the 17th International Conference on Artificial Intelligence in Education (in press)
13. Segedy, J., Biswas, G., Blackstock, E., Jenkins, A.: Guided skill practice as an adaptive scaffolding strategy in open-ended learning environments. In: Proceedings of the 16th International Conference on Artificial Intelligence in Education. pp. 532-541. Springer (2013)