Scanpath Analysis of Student Attention During Problem Solving with Worked Examples

Samantha Stranc and Kasia Muldner(✉)

Institute of Cognitive Science, Carleton University, Ottawa, Canada
{samantha.stranc,kasia.muldner}@carleton.ca

Abstract. We report on the analysis of scanpath data captured by an eye tracker as students solved problems with access to worked examples. Our work makes two contributions: (1) it reports on scanpath analysis using the MultiMatch tool, (2) it investigates how type of problem-example similarity and assistance influenced attention patterns captured by scanpaths. We show that both problem-example similarity and type of assistance impact scanpaths.

Keywords: Scanpath analysis · Problem solving & worked examples · Assistance

1 Introduction

Eye tracking data can provide valuable data on students' visual attention, which can then be used as input to a user model to detect various student states [1–3]. To date, analysis has focused on *fixation* data [4], namely a moment of attention when the eye stops scanning. Additional insight can be gained by accounting for the *order* of fixations, captured by a *scanpath* [5]. For instance, sequence mining can identify common patterns in scanpaths [3, 6]. A limitation of this method is that it only considers exact sequence matches. However, two scanpaths are rarely identical and this is particularly the case for longer scanpaths, such as ones elicited by complex tasks. Consequently, the results from sequence mining are typically abbreviated to only include sequences consisting of 2 to 4 fixations. This has the potential to miss information on strategies involving longer sequences of fixations. We address this limitation by using fuzzy alignment approaches provided by a scanpath tool called MultiMatch [7]. MultiMatch transforms the original coordinate data into a vector-space representation and then quantifies the similarity between two scanpaths using five features (shape, direction, length, position and duration). Here, we use MultiMatch to analyze the similarity of scanpaths captured as students solved algebra problems in the presence of examples using a basic computer tutor.

Scanpath analysis has been applied in a range of domains like scene analysis [8], decision making [9], reading [10], and analogy making [11]. For instance, Zhou et al. [9] analyzed scanpaths in three different decision-making tasks (e.g., one task involved choosing between risky options under two different conditions). The similarity of scanpaths in a given decision condition were more similar than between the conditions, suggesting that visual attention was affected by type of task. However, more work is

I. I. Bittencourt et al. (Eds.): AIED 2020, LNAI 12164, pp. 306–311, 2020.
https://doi.org/10.1007/978-3-030-52240-7_56

needed to investigate the potential utility of scanpath analysis for educational contexts, something echoed in recent reviews [12].

2 Data and Methods

The instructional context for our work was problem solving with access to worked examples. In this context, learning outcomes depend on student *strategies*, including how much students copy from examples (bad for learning) vs. self-explain from examples or problems (good for learning) [13–15]. To date, these strategies have been investigated using analyses of student utterances [13] and so less is known about students' visual attention in this context. Here, we used data from a between-subjects eye tracking study [16] in which students used a tutoring system to solve 12 algebra problems, with assistance from one example per problem. Half of the problem-example pairs had *high similarity* (example solution could be copied), while the other half had *low similarity* (inferences beyond copying were required to apply the example). Here, we focus on two study conditions: (1) *fade-out assistance* (n = 20): students were initially given high-similarity examples, but these transitioned to low similarity after some problems were solved; (2) *fade-in assistance* (n = 19): the opposite was the case (low-similarity initially, eventually becoming high similarity). Thus, while the conditions involved the same number of low and high similarity examples, the *timing* of assistance was varied (immediate presentation of high similarity examples, vs. later in the problem sequence).

The original analysis showed that students learned more from *fade-in* assistance than *fade-out* assistance [16] but did not analyze students' strategies. Here, we analyze the scanpaths in each condition (*fade in* vs. *fade out*), using them as a proxy for strategies. Recall that scanpaths are series of fixations on learning materials. If we can show that scanpaths are different, this provides some evidence that strategies are different as well.

We had two key questions: (1) Does problem-example similarity impact scanpaths? (2) Does the type of assistance (fade in vs. fade out) impact scanpaths? Recall that there were 12 problems solved in each condition (*fade in* and *fade out*). We analyzed scanpaths from the 1st and the 9th problem-example pair. The 1st and 9th problem were paired with a high-similarity example in the fade-out condition and a low-similarity example in the fade-in condition. Thus, comparing scanpaths from the 1st problem in each condition allowed us to analyze the effect of problem-example similarity, before any impact of condition took place. Since the two conditions involved a different structuring of assistance (fade in vs. fade out), analyzing scanpaths from the 9th problem-example pair, with the analysis from the 1st pair serving as the baseline, allowed us to investigate the impact of assistance on how problem 9 was solved.

Method. We followed the standard method [9, 17] to analyze the data. We extracted the scanpaths for each problem-example pair per participant in each condition (we capped the scanpath length at 500 fixations to make the analysis feasible; the majority of scanpaths were shorter). We then used MultiMatch to compare all possible pairs of scanpaths (a) *within* a given condition; (b) *between* the two conditions. Each comparison produces one similarity score for each of the 5 MultiMatch features. If students are using a different strategy in the *fade-in* and *fade-out* conditions, then the average scanpath similarity score from *fade-in* condition should be different from *fade-out* condition. However, if the similarity scores are about the same for the *fade-in* and *fade-out* conditions, this shows students are consistent in processing the material *within* each condition, but we don't know if they are using the same strategy in the two conditions. In this case, we also need to compare *between* conditions to obtain a benchmark for the within analyses (for details on this methodology, see [9, 17]). The final step involved analyzing the data using inferential statistics. This analysis is at the *similarity-score* level and so does not violate the independence assumption because each data point corresponds to a participant_i-participant_j score that only appears once in the overall analysis (this strategy was used in prior work [17]).

3 Results and Discussion

Recall that MultiMatch produces 5 similarity scores per scanpath-pair comparison, one per feature (shape, direction, length position, duration). The similarity scores range between 0 and 1 (higher = more similar). MultiMatch produces high scores [17] and so the relative difference in scores is more informative than raw scores. We had 3 groups, based on similarity scores from comparing scanpaths *within* the fade-in condition ($\text{Sim}_{\text{fade in}}$), *within* the fade-out condition ($\text{Sim}_{\text{fade out}}$) and *between* the two conditions ($\text{Sim}_{\text{between}}$). Thus, we used an ANOVA with *comparison group* as the 3-level factor.

Analysis 1. Analysis 1 focused on scanpaths extracted from the first problem-example pair in each condition. The descriptives for the main statistics are in Table 1. The ANOVA on the similarity scores reported significant results for two MultiMatch features: direction, $F(2, 663) = 10.11$, $p < 0.01$, and shape, $F(2, 663) = 50.72$, $p < 0.01$. For direction, pairwise comparisons showed that $\text{Sim}_{\text{fade out}} > \text{Sim}_{\text{fade in}}$ ($p < .01$). Since for problem 1, the *fade-out* group had a high similarity problem-example pair and the *fade-in* group had a low similarity pair, this result shows that scanpaths are more similar when students are given high-similarity examples. While for shape there was also a significant effect of condition, the raw effect size was small. Moreover, scanpaths were significantly different between the conditions, showing that problem-example similarity influenced how students viewed the problem and the example. (The two pairwise comparisons involving $\text{Sim}_{\text{between}}$ were also significant, but these are not as informative given the significant difference between $\text{Sim}_{\text{fade in}}$ and $\text{Sim}_{\text{fade out}}$).

Analysis 2. The second analysis focused on problem 9. By the time problem 9 was encountered, participants had experienced the effect of assistance type (fade in vs. fade out). Importantly, at problem 9, the corresponding example had the same similarity as for problem 1, and so problem 1 served as a baseline. Descriptives are in Table 1. Because we used problem 1 as the baseline, we ran the analysis on the similarity difference scores (problem 9 – problem 1). The ANOVA reported significant results for 3 MultiMatch features: length, $F(2, 627) = 4.8$, $p < 0.01$, position, $F(2, 627) = 10.8$, $p < 0.01$, and duration, $F(2, 627) = 3.5$, $p = 0.03$. Pairwise comparisons for each

Table 1. Descriptives (mean, stDev) for MultiMatch features. For problems 1 and 9, assistance was low for the fade-in group and high for the fade-out group.

	Shape	Direction	Length	Position	Duration
Problem 1					
fade in	.9873 (.002)	.7506 (.066)	.9851 (.003)	.9163 (.039)	.6627 (.032)
fade out	.9847 (.003)	.7765 (.037)	.9843 (.003)	.9107 (.038)	.6676 (.032)
Problem 9					
fade in	.9866 (.002)	.7374 (.126)	.9842 (.004)	.9181 (.035)	.6738 (.040)
fade out	.9840 (.003)	.7865 (.043)	.9818 (.005)	.8947 (.046)	.6721 (.037)

feature were significant ($p < .01$). For length and position, the difference in similarity scores from problem 1 to problem 9 was significantly greater for the *fade-out* group. This is reflected in the descriptives: for the fade-out group, the average similarity score is bigger for problem 9 than problem 1 for these features, while for the fade-in group, the average similarity score is virtually identical for problem 1 and problem 9. The duration feature, however, demonstrated the opposite pattern with significantly higher change in similarity scores for the *fade-in* group than the *fade-out* group.

To summarize, we found that (1) problem-example similarity affects scanpaths (analysis 1); and type of assistance affects scanpaths (analysis 2). As far as analysis 1, our results confirm prior work showing that difficulty reduces scanpath similarity [17]. The low similarity problem-example pair in the *fade-in* condition was more difficult because it blocked copying of the example (an easy strategy) and so required problem solving (a harder strategy). Analysis 2 examined how assistance influenced change in scanpaths. When the problem-solving session started with low similarity problem-example pairs (*fade-in* group), students were blocked from copying the example solutions. The original analysis [16] found that students subsequently viewed the problem more than the *fade-out* group. Thus, the fade-in group's strategy corresponded to attention to the problem - the present analysis suggests this strategy remained stable over time for the length and position features (there was little difference in scanpaths between problem 1 and 9 for these features). In contrast, for these features, the *fade-out* group's scanpaths changed over time (similarity higher at problem 9 than problem 1), suggesting this group may have started out with one strategy (copying from the high similarity examples they initially received) but revising this strategy when assistance faded out.

Our results show effects for distinct MultiMatch features. Space constraints prevent us from in-depth discussion, but we offer brief interpretations. For analysis 1, direction was one of the informative features. Similarity should be low for this feature when saccades are moving in opposite directions from each other in the target scanpaths. This was occurring more in the *fade-in* group who had the low-similarity example, as the similarity for direction was lower than the *fade-out* group who had the high-similarity example. Thus, type of example influenced the direction of saccades. Additionally, we speculate that length and position, significant for analysis 2, relate to the location of the gaze. Since the *fade-out* group had reduced similarity for these features on problem 9 compared to problem 1, we can speculate they changed their strategy in terms of what they looked at. The implication of our work is that scanpath analysis can identify differences in visual attention for problem-solving tasks. However, more work is needed to identify the benefits and limitations of a scanpath approach.

Acknowledgements. This work was supported with an NSERC Discovery Grant 1507.

References

1. Hutt, S., Mills, C., White, S., Donnelly, P.J., D'Mello, S.: The eyes have it: gaze-based detection of mind wandering during learning with an intelligent tutoring system. In: Proceedings of the 9th International Conference on Educational Data Mining (EDM 2016), pp. 86–93 (2016)
2. Muldner, K., Burleson, B., VanLehn, K.: "Yes!": using tutor and sensor data to predict moments of delight during instructional activities. In: Proceedings of User Modeling, Adaptation and Personalization Conference, pp. 159–170 (2010)
3. Taub, M., Azevedo, R.: How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive SRL processes during learning with an intelligent tutoring system? Int. J. Artif. Intell. Educ. **29**(1), 1–28 (2019)
4. Lai, M.L., Tsai, M., et al.: A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ. Res. Rev. **10**, 90–115 (2013)
5. Anderson, N.C., Anderson, F., Kingstone, A., et al.: A comparison of scanpath comparison methods. Behav. Res. **47**, 1377–1392 (2015)
6. Steichen, B., Wu, M., Toker, D., Conati C., Carenini, G.: Te, Te, Hi, Hi: eye gaze sequence analysis for informing user-adaptive information visualizations. In: Proceedings of UMAP 2014, pp. 133–144 (2014)
7. Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., Holmqvist, K.: It depends on how you look at it: scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. Behav. Res. Methods **44**(4), 1079–1100 (2012)
8. Foulsham, T., Underwood, G.: What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. J. Vis. **8**(2), 6 (2008)
9. Zhou, L., Zhang, Y., et al.: A scanpath analysis of the risky decision-making process. J. Behav. Dec. Making **29**(2–3), 169–182 (2016)
10. Von der Malsburg, T., Angele, B.: False positives and other statistical errors in standard analyses of eye movements in reading. J. Mem. Lang. **94**, 119–133 (2017)
11. French, R., Glady, Y., Thibaut, J.: An evaluation of scanpath-comparison and machine-learning classification algorithms used to study the dynamics of analogy making. Behav. Res. Methods **49**(4), 1291–1302 (2017)
12. Lai, M.L., et al.: A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educ. Res. Rev. **10**, 90–115 (2013)
13. Chi, M.T.H., Bassok, M., Lewis, M., Reimann, P., Glaser, R.: Self- explanations: How students study and use examples in learning to solve problems. Cogn. Sci. **13**, 145–182 (1989)
14. VanLehn, K.: Analogy events: how examples are used during problem solving. Cogn. Sci. **22**(3), 347–388 (1998)
15. VanLehn, K.: Rule-learning events in the acquisition of a complex skill: an evaluation of cascade. J. Learn. Sci. **8**(1), 71–125 (1999)

16. Jennings, J., Muldner, K.: From dissimilar to similar: reverse fading assistance improves learning. In: Proceedings of the Cognitive Science Society, pp. 560–566 (2018)
17. Dewhurst, R., Foulsham, T., Jarodzka, H., Johansson, R., Nyström, M.: How task demands influence scanpath similarity in a sequential number-search task. Vision. Res. **149**, 9–23 (2018)