Measuring Navigation Performance in Serious Games

Chris Ferguson, Herre van Oostendorp, Geert-Jan Giezeman, Sam de Redelijkheid, Egon L. van den Broek

Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands.
c.ferguson@uu.nl

Abstract: In serious games, in-game analytics is a major focus, as it is beneficial to see how a player is learning throughout the game. Different control schemes, such as ones often used in Virtual Reality (VR) games, open the door to a variety of existing analytical techniques. This article compares three measures adopted from two distinct areas to determine player's navigation behaviour, namely: task-finding lostness, gathering lostness and sequence similarity. Results on 13 children playing a serious Virtual Reality (VR) game, show both resemblances and differences among the three measures. Moreover, each of the measures show their constraints when applied in a serious VR game context. If anything, the current article illustrates, once more, how complex the analysis of player behaviour is and the need for the absent ground truth measure.

Introduction

Particularly in serious games, in-game analytics is a major focus, as it is beneficial to see how a player is performing in the game. Different control schemes, such as ones often used in Virtual Reality (VR) games, namely node-based movement, open the door to a variety of existing analytical techniques. Some of which can be adopted from other areas of computing, that could possibly be used to evaluate a player's performance within the game [11,12]. These are different to dedicated in-game analytics, which usually focus on heatmaps [1] or identifying switching behaviours, and the number of game objects (e.g., videos, resources, and locations) accessed per time unit [2].

This article compares three measures from two different areas of computing to determine player's (dis)orientation: task-finding lostness, gathering lostness and similarity sequence. The two lostness measures were originally introduced to analyse navigation in hypertext systems [5]. However, we hypothesize that these can be used for serious game navigation as well, as navigation is a key issue in the learning process. Additionally, a sequence similarity measure is included [6] that has been successful in a range of distinct domains including bioinformatics and music analysis. As such, this article contributes to methods and techniques in game playing analytics.

Distance measures

Lostness

The lostness measure [4] was created as a method for measuring hypertext usability in terms of task performance. The starting point for this was the belief of the authors that 'measures based on time and errors seem inappropriate for hypertext systems which, by their nature, encourage exploration and browsing'. Instead, a better approach is to assess task performance in terms of the effectiveness in how users find information and the degree in which they become lost in the information search whilst looking for this information [5].

Lostness [4] is defined by the number of information items inspected compared with the number of items that need to be inspected to locate the requested information:

$$\sqrt{\left(\frac{N}{S}-1\right)^2+\left(\frac{R}{N}-1\right)^2},$$

with R, S, and N being respectively the minimum number of nodes that need to be visited to complete a task, the total number of nodes visited whilst searching, and the number of different nodes visited whilst searching.

Table 1. Comparison between old scope and new scope.

Initial (Hypertext) Scope	New (Game) Scope
Clicking a link to move to a new web page	Moving to another node within the environment
Inspecting a web page containing information	Picking up an item
Finding information on a web page	Examining an item

R.A Grant et al. (Eds.): Measuring Behavior 2018, ISBN 978-1-910029-39-8
Manchester, UK, 5th-8th June 2018

Gathering versus Fact-Finding Lostness
Based on the identification of two types of activities, gathering and fact-finding, two lostness measures were constructed based on these. The gathering lostness measure involves calculating the minimum steps necessary to complete the whole task whereas fact-finding lostness calculates the lostness between each of the objectives in the task.

The difference between these two measures is that the gathering lostness is based on the number of steps identified manually imputed based on watching a perfect playthrough but, contrarily, fact-finding lostness is more automated. Instead, the spatial graph created for node-based movement, which defines which movement nodes are accessible from each node, is traversed and distance between each objective and the one accessed previously is calculated using a breath first search and is matched to the players path. This provides the minimum number of steps. This gives the opportunity to investigate if lostness is more representative if it is calculated per each objective or the whole task.

Needleman-Wunsch Similarity
Musaline [6] is a C++ library for alignment of sequences, which was initially developed for use with musical sequences; but, can also be used for comparing other sequences. It makes use of the Needleman-Wunsch algorithm [7]. In contrast to the lostness measure, 100% shows a perfect match, whereas 0% shows no match at all. To figure out how lost a player is, the path followed by the player is compared with the optimal path. While the lostness measure looks at path length and the number of unique nodes visited, the Needleman-Wunsch algorithm looks at two other aspects, namely:

- Identity, the visited nodes should be identical to nodes in the optimal path
- Order, the nodes should be visited in the same order as in the optimal path.

Given an optimal path and a player path, the Needleman-Wunsch algorithm efficiently searches through all possible alignments between both paths. An alignment is a list of nodes, where each node represents either a node in the optimal path, the player path or both, such that for each alignment we can reconstruct the user path by taking all nodes representing the user path and the optimal path by taking all nodes representing the optimal path.

A node in an alignment is considered correct if it represents both the player path and the optimal path, otherwise it is considered incorrect. We assign a cost to all incorrect steps and assigning a reward to each correct step. The similarity of an alignment is defined by summing over the costs and rewards for each step. Finally, the similarity of the paths is defined by taking the maximum similarity of all.

The cost and reward function chosen for the Chantry was based on several different user paths in the game and is 1 for a correct step and -0.35 for an incorrect step. Then we divide the similarity value by the number of steps in the optimal path and clamp it at 0 to create a value between 0 and 1. The values are chosen to match the gathering similarity values as closely as possible.

Weighted Measures
Simply taking the average of the lostness or similarity values does not suffice to obtain an exact overview of the player performance over the full game. The relatively short paths can be prone to noise and, due to the diminishing effect of extra steps, a player that gets completely lost on one or two locations may have a lower average lostness than someone who never gets completely lost; but, instead takes a side step on some otherwise perfect paths. Conversely, player performance could be overstated as low levels of lostness on simpler tasks may mask higher levels of lostness in more complicated tasks. Therefore, the complexity of tasks must be taken into account and weighted.

Although, for gathering lostness and sequence similarity, the values are simply weighted depending on the number of objectives needed to be completed in each gathering task, a different approach is taken for the fact-finding lostness. As well as calculating the lostness value for each objective, as the player also needs to move from the end of the objective to the start of another, the steps between the objectives are also taken into consideration. Therefore, this provides a weighted fact-finding lostness measure for the whole game:

$$\sqrt{\left(\frac{\sum_{o=1}^{n} N_o}{\sum_{o=1}^{n} S_o}-1\right)^2 + \left(\frac{\sum_{o=1}^{n} R_o}{\sum_{o=1}^{n} N_o}-1\right)^2},$$

with n being the number of consecutive objective pairs in the player's path through the game and R_o, N_o, and S_o are respectively the required steps, total steps taken, and unique steps taken for the o^{th} objective pair.

Method

A total of 13 children (6 boys) aged 13-18, were recruited. Their VR experience varied substantially.

A PlayStation VR headset and over-ear headphones were used to secure the player's immersion with the game. A PlayStation controller was used to interact with the game.

Game play behavior was recorded time stamped, in the background. Consequently, it was possible to analyse the paths taken by the players and, hence, calculate the lostness and sequence similarity measures. Participants were given 30 minutes to play through the game at their own pace, learning about the story and completing tasks.

"The Chantry", an educational environmental narrative game for the PlayStation VR platform, was used as testbed. It tells the story of dr Edward Jenner and his invention of vaccination against the smallpox virus. To progress through the game, the player needs to explore the house of Jenner, finding out information about a particular story topic before moving on. Players interact with objectives that contain story information that must be found in order to continue. The full task is best described as a 'gathering activity' as the target information is spread out over different areas and has to be combined, whereas, finding each of the objectives present in the list can be referred to as more simple, local 'fact-finding activities' as all of the information is located in a specific place [3].

Results

The weighted lostness and sequence similarity values were investigated for all participants, to see how well they match each other. For this, the inverse of the lostness measures were used so that the perfect score represented by both measures would be 100%. So, a high (inverse) lostness score means that the participant exhibits a perfect search. The outcomes are reported in Table 2.

Overall, participants were not overly effective at the tasks, with all measures being below 54%. The difference between the gathering lostness and sequence similarity means is quite small (1.67), though fact-finding lostness is notably higher than both of these measures (9.47 higher than gathering lostness and 11.14 higher than sequence similarity). Moreover, compared to both the inverse gathering lostness and sequence similarity, the fact-finding measure gives somewhat different results. Inverse gathering lostness and sequence similarity (Pearson correlation .93, $p < .0001$) are almost equivalent, which was expected as the cost function of the sequence similarity was fitted to match the lostness value.

The data show that, across all measures, males performed better than females. This could be explained by males having more VR experience (mean: 2.33) compared to the females (mean: 1.71). However, these results could be contributed to gender differences in visual-spatial ability [8,9].

Table 2. Results for both inverse (weighted) lostness and the sequence similarity measures (in %).

	Gathering Lostness	Fact-Finding Lostness	Sequence Similarity
Total mean (SD)	44.07 (19.13)	53.54 (11.19)	42.40 (19.07)
Female mean (SD)	41.19 (22.38)	49.12 (10.63)	41.24 (23.43)
Male mean (SD)	47.42 (15.89)	58.69 (10.26)	43.74 (14.49)

Discussion

Altogether, three known analytic measures were introduced to a new domain: serious VR games. The measures indicate similar but, in parallel, distinct player behaviour. Additional studies must be carried out in order to identify which of these lostness measures is more accurate, which may also involve a fact-finding sequence similarity measure.

We signalled that each of the three measures has its constraints. For the lostness measures, the minimum steps necessary to complete a task must be constant. If the minimum value can unexpectedly change; then, an incorrect value will be returned. In addition, the path for lostness must be completely linear. Due to the nature of the measure, which is designed to punish node revisits, even if the minimum steps are correct, the user will be identified as lost. This is shown in the current study where one of the tasks returns a lostness value of 0.32, even when it has been completed perfectly. The sequence similarity allows multiple optimal orders, which is not always wanted. This is shown in the current study, where most tasks had 1-4 perfect paths, going up to 18 in one case.

One of the most illuminating results was that the mean task performance over most lostness measures was quite low, indicating a high level of disorientation in participants. This backs up that, as games are complex environments, players need to learn how to control the game and how it conveys the instructional material, before this material can be learned [10]. Therefore, to improve task performance, players should play a short tutorial to familiarise themselves with the environment, gameplay controls and mechanics, particularly the node movement system and picking up and using items.

Three distance measures were introduced to the domain of serious VR games. Each showed its behaviour and own constraints. If anything, they illustrated how complex the analysis of player behaviour is and the need for the absent ground truth measure. Results indicate the need for both basic research on intuitive distance measures in line with human behavior and their validation in ecological valid settings, such as serious VR games.

Acknowledgments
This work was supported by the REVEAL (Realising Education through Virtual Environments and Augmented Locations) Project, project number: 732599, funded under the European's Union's Horizon 2020 Research and Innovation Programme. Innovative Action (IA) H2020-ICT-2016-1. The experiment performed was carried out on the grounds of Sheffield Hallam University and approved by the University Ethics Committee (SHUREC).

References

1. Drachen, A., Schubert, M. (2013). Spatial Game Analytics. In: Seif El-Nasr M., Drachen A., Canossa A. (Eds.) *Game Analytics*. London, UK: Springer.
2. Westera, W., Nadolski, R., Hummel, H. (2014). Learning Analytics in Serious Gaming: Uncovering the Hidden Treasury of Game Log Files. In: A. de Gloria (Ed.), *Games and Learning Alliance. GALA 2013*. Lecture Notes in Computer Science, vol. 8605. Cham, Switzerland: Springer.
3. Puerta Melguizo, M.C., Vidya, U., van Oostendorp, H. (2012). Seeking information online: the influence of menu type, navigation path complexity and spatial ability on information gathering tasks. *Behaviour & Information Technology* **31**(1), 59-70.
4. Smith, P. (1996). Towards a practical measure of hypertext usability. *Interacting with Computers* **8**(4), 365-381.
5. Gwizdka, J., Spence, I. (2007). Implicit measures of lostness and success in web navigation. *Proceedings of the 31th European Conference on Cognitive Ergonomics (ECCE2013),* Article no 17. New York: ACM.
6. Utrecht University (2012). Musaline: Musaline library. [online] Projects.science.uu.nl. Available at: https://www.projects.science.uu.nl/monochord/musaline/index.html [Last accessed 24 Apr. 2018].
7. Needleman, S.B., Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**(3), 443-453.
8. Meijer, F., van den Broek, E.L. (2010). Representing 3D virtual objects: Evidence for an interaction between visuo-spatial ability and the nature of exploration. *Vision Research* **50**(6), 600–635.
9. Voyer, D., Voyer, S., Bryden, M.P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. Psychological Bulletin **117**(2), 250-270.
10. Wouters, P., van Oostendorp, H. (2013). A Meta-analytic Review of the Role of Instructional Support in Game-based Learning. *Computers & Education* **60**(1), 412-425.
11. van den Broek, E.L., Kuijpers, R. (2014). Game Over! On measurement and optimization of presence. In A.J. Spink, E.L. van den Broek, L.W.S. Loijens, M. Woloszynowska-Fraser, L.P.J.J. Noldus (Eds.), *Proceedings of Measuring Behavior 2014: 9th International Conference on Methods and Techniques in Behavioral Research*, p. 43–47. August 27–29, Wageningen, The Netherlands.
12. Sharma, K., Castellini, C., Stulp, F., van den Broek, E.L. (2018). Continuous, real-time emotion annotation: A novel joystick-based analysis framework. *IEEE Transactions on Affective Computing* **9**. DOI: 10.1109/TAFFC.2017.2772882. [in press]