

A Connectionist Model of Spatial Knowledge Acquisition in a Virtual Environment

Corina Sas¹, Ronan Reilly² and Gregory O'Hare¹

¹Computer Science Department, University College Dublin
Belfield, Dublin 4, +353 1 716 {2922, 2472}
{corina.sas, gregory.ohare}@ucd.ie

²Computer Science Department, National University of Ireland
Maynooth, Co. Kildare, +353 1 708 3846
ronan.reilly@may.ie

Abstract. This paper proposes the use of neural networks as a tool for studying navigation within virtual worlds. Results indicate that the network learned to predict the next step for a given trajectory, acquiring also basic spatial knowledge in terms of landmarks and configuration of spatial layout. In addition, the network built a spatial representation of the virtual world, rather like a cognitive map. The benefits of this approach and the possibility of extending the methodology to the study of navigation in Human Computer Interaction applications are discussed.

1 Introduction

This study is part of an ongoing research the purpose of which is to identify the procedural and strategic rules governing user navigational behavior within virtual space. Environmental psychology provides an empirical basis for a better understanding of how humans perceive and understand the *space*. The work described here confirms the idea that acquiring an internal representation of the environment is a very complex process involving primarily landmarks identification and understanding of spatial layout configuration. These two basic procedures are well known as *route-based knowledge* and *survey knowledge* [8]. Without underestimating the role of traditional methods, we propose the use of neural networks as an alternative tool for studying navigation within virtual worlds.

Neural networks have proven particularly suited to finding patterns in large amounts of complicated and imprecise data, and detecting trends that are too complex to be noticed by humans [2]. While neural networks have been fruitfully exploited by artificial intelligence researchers, their adoption within HCI has been limited. They have been primarily applied to pattern recognition [18]. Finlay identified four areas of HCI, which involve pattern recognition problems, such as task analysis and task evaluation, natural interaction methods such as gesture, speech, handwriting, and adaptive interfaces [5]. Neural networks provide a very powerful toolbox for modeling complex high-dimensional non-linear processes [13]. ANNs have many advantages over the traditional representational models, in particularly distributed

representations, parallel processing, robustness to noise or degradation and biological plausibility [10]. We consider that at least part of these strengths can be harnessed to model user's behavior in terms of spatial knowledge acquisition. A simple recurrent network [4] has feedback which embodies short-term memory. This makes it suitable for application to symbolic tasks that have a sequential nature.

2 Navigation within Virtual Environments

Virtual Environments (VE) have become a rich and fertile arena for investigating spatial knowledge. Within the VE, the user's set of *actions* is restricted, consisting mainly of navigation, locomotion, object selection, manipulation, modification and query [6]. As Sayers (2000) observed navigation has been found to be central to the usability of interfaces to VEs on desktop systems [21]. VEs offer the context for training and exploration, enabling the replacement of training and exploration within the physical world. This proves partially attractive when experiencing the real world is expensive, dangerous or hard to be achieved [3]. Evidence of significant similarities in the acquisition of spatial knowledge from real and VEs has been identified [11]. A further advantage consists of their powerful tractable characteristic [1], which enables accurate spatio-temporal recording of users' trajectory within the virtual space. Attempts to understand spatial behavior in both real and artificial worlds were primarily concerned with highlighting the symbolic representation of spatial knowledge.

2.1 Symbolic Cognitive Models of Navigation

The study of navigation in the area of HCI has developed mostly in the field of cognitive modeling, benefiting from inputs provided by both environmental psychology and geography [14]. Modeling of spatial knowledge has constituted a central research theme for the last four decades. Golledge presented different models of declarative knowledge acquisition, together with their relevant applications in the area of spatial cognition [7]. Kuipers developed several computational models for navigation, underlying the procedural knowledge embedded in the spatial representations [12]. The basic idea resides in the individuals' set of interactions with the environment, which facilitates a structured storage of perceptual experiences. These memorized experiences would enable users to build a more generalized structure for exhibiting an emergent spatial behavior unperformed before [7].

2.2 Connectionist Models of Navigation

Previous studies have shown that recurrent neural network can predict both circular and figure eight trajectories [4,9,17,19,22]. However, due to the fact that the figure eight trajectory crosses itself, the training was more difficult for this type of trajectory. In our case, the trajectories covered by users are more complex than a circle or figure eight, even though some of them resemble a circular shape.

3 Methodology

Research in the area of navigation within VEs has been generally focused on large-scale virtual worlds [3]. In this study we utilized ECHOES¹ [16], as an experimental test-bed. It is a virtual reality system, which offers a small-scale world, dense, static and with a consistent structure. Adopting a physical world metaphor, the ECHOES environment comprises a virtual multi-story building, each one of the levels containing several rooms: conference room (Fig.1), library (Fig.2), lobby etc.

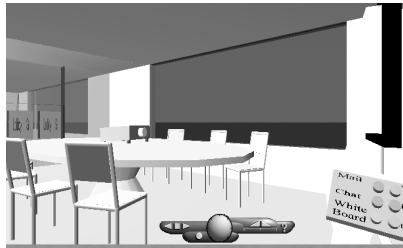


Fig. 1. Virtual Conference Room



Fig. 2. Virtual Library

The present study captures the spatial behavior of users exploring an unfamiliar VE. Users can navigate from level to level using a virtual elevator. The rooms are furnished and associated with each room are a set of user functions.

A sample of 30 postgraduates in the Computer Science Department was asked to perform two tasks within the virtual world, namely exploration and searching. In order to gain familiarity with the environment and learn movement control, the subjects were asked to look for a particular object within the virtual building for about 20 minutes. This exploratory task provided the primary data for the neural network approach. Furthermore, subjects were asked to find a particular room in the virtual building, namely the library. We considered the time and length of trajectory involved in this searching task as *performance indicators*. Based on these, we identified the quality of spatial knowledge acquisition and the efficiency of the exploratory strategy.

A comprehensive set of data consisting of users' positions was recorded throughout the experiment. Each movement greater than half a virtual meter, and each turn greater than 30° were recorded.

We present a connectionist simulation to test whether a network can build a *cognitive map* as an internal representation of environmental information [8] in terms of both landmarks and configuration of the spatial layout. The basic idea is that an input vector consisting of current Cartesian coordinates together with information about the nearest landmark is sufficient to induce the network's internal abstractions for predicting the next position. To test our hypothesis, an Elman-style simple recurrent neural network was used to learn the trajectory and to predict the next step.

¹ ECHOES (European Project Number MM1006) is partially funded by the Information Technologies, Telematics Application and Leonardo da Vinci programmes in the framework of Educational Multimedia Task Force.

The implementation was carried out by using Stuttgart Neural Network Simulator (SNNS). The network architecture [4] is presented in Figure 3 and consists of 6 input nodes, 12 hidden nodes, 12 context nodes and 6 output nodes.

The network input consists of a sequence of users' trajectories. At each time step t , an input vector is presented consisting of user's position, orientation angle, distance to the nearest landmark (the distance to the nearest point of the landmark) and its associated position (coordinates of the center of the landmark). For this simulation we considered only the trajectories performed on the ground floor of the virtual building. Figure 4 presents an overhead image of this level. After each trajectory was entered, the network has been reset, by entering an input which zeros out the output [15]. The output pattern represents the input vector of time $t+1$. All the input values have been normalized.

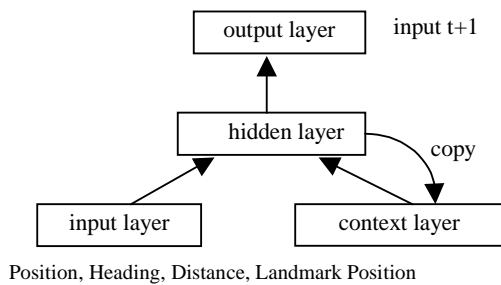


Fig. 3. Network Architecture

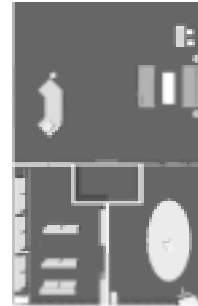


Fig. 4. Ground Floor

Using the backpropagation learning procedure [4] the network was taught to predict for each current position the next position in time. Table 1 presents an example of input/output vectors, which reflect a rotation (the third element in the vector changes), followed by a translation which moves the user in the proximity of a new landmark (the last two elements in the vector change).

Vector	X	Y	Rotation	Distance	Landmark (X)	Landmark (Y)
Input t	0.031	0.258	0.097	0.179	0.000	0.078
Output t	0.031	0.258	0.018	0.179	0.000	0.078
Input $t+1$	0.031	0.258	0.018	0.176	0.000	0.078
Output $t+1$	0.025	0.336	0.018	0.195	-0.219	0.531

Table 1. Input / Output Vectors Example

At this stage of our work, we expanded the notion of landmark to any feature added to spatial layout. Therefore, apart from any piece of furniture, we considered also the choice points such as doors and lift entrance. Identifying which ones among these features prove to be salient and able to capture attention – being thus an authentic landmark – is a task to be solved by the network.

We divided randomly the entire set of data into five parts, using three of them for training, one for validation and one for testing. The network was trained for 1000

epochs, with 24 trajectories composed of 4668 input vectors. It was tested with 12 trajectories consisting of 1573 input vectors. The average trajectory length was 160 vectors. The learning rate was 0.001, the initial weights set within a range of 0.5 and the momentum was 0.

4 Results

In Table 2 we present the results of testing the network, obtained by computing the Euclidean distance between the output vectors predicted by the network and the expected output vectors.

Input description	Percent Correct
User's next position – X coordinate	97.13 %
User's next position – Y coordinate	92.30 %
User's next orientation (heading)	86.90 %
Distance to the next nearest landmark	99.87 %
Next nearest landmark position – X coordinate	90.27 %
Next nearest landmark position – Y coordinate	86.77 %

Table 2. Prediction accuracy of each input element based on Euclidean distance

As can be seen, the network generalizes well for all the input elements. However, for a prediction to be correct all the input elements should simultaneously be within specified limits (e.g. 1 virtual meter for Cartesian coordinates, 30 degree for rotation and 2.5 virtual meters for distance estimation). With respect to the composite criterion of accuracy, the network still performs adequately, the success rate being 67.57%.

5 Discussion

The preliminary results obtained by training the recurrent neural network proved promising, indicating that the network not only learns to predict the next step for a given trajectory, but it also learns the spatial layout in terms of landmark configuration.

In order to highlight the internal representation of the network, we performed a series of analyses. Firstly, we were concerned whether the network could learn any boundaries. Since all the vectors predicted were within the limits delineating the ground floor layout, it seems that the network did indeed learn the floor boundaries. Another issue we were interested in was whether the network built any cognitive map of the virtual space.

Figure 5 presents the “cognitive map” derived from network representation of landmarks’ positions. It shows the layout of the virtual space (ground floor) in terms of the centers of the landmarks, e.g. (x, y) coordinates, as the neural network has predicted them. As it is presented, the predicted coordinates of the landmarks form clusters around the coordinates of the landmarks within the VE. The centroid for each

cluster is plotted as well. The “cognitive map” (the internal representation of space achieved by the network) conserves well the topology while its metric is slightly less accurate. The same properties characterize the cognitive maps built by humans. In other words the network built a map of the space in a way similar to humans.

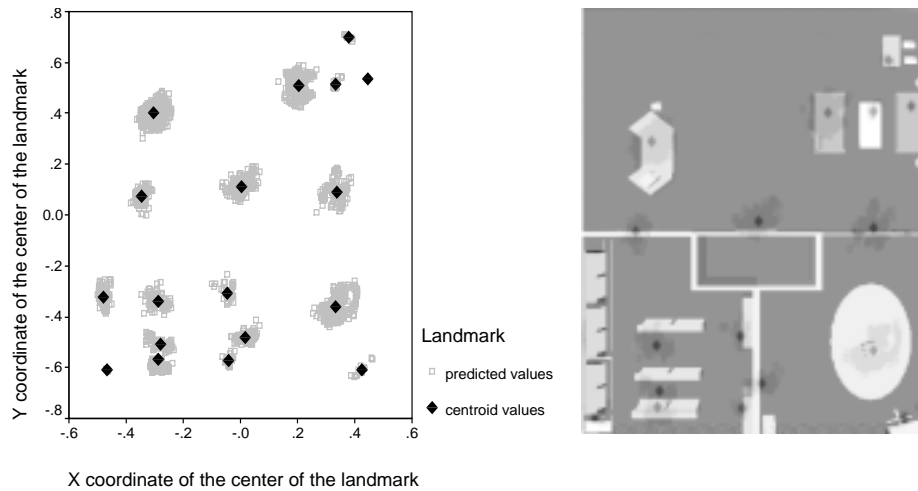


Fig. 5 Centers of the landmarks predicted by the neural network

For a better understanding of the network’s ability to discriminate between landmark features, we performed an analysis of network predictions regarding the attention it paid to each landmark. More precisely, we counted how many times a landmark was visited, or in other words how many times a given landmark was the nearest to the user. We took this measurement as an indicator of landmark saliency.

The most important landmarks are the desk with a computer, the sofa in the center of the larger room, the door to the meeting room, the large elliptical table, the lift and the door between meeting room and library.

The saliency of a landmark is related to the landmark’s location in the room, e.g. centrality, its size, and unique features (e.g. shelves in the library are all alike thus undistinguishable). These findings suggest that users paid particular attention to connectivity/decision points such as doors and the lift.

For understanding the network accuracy of predicting the landmarks, the figure 6 presents the actual centers of the landmarks together with the centroids of the centers of the landmarks, as the neural networks anticipated them.

A visual examination suggests that the network predictions of landmarks locations, in terms of (x, y) coordinates, approximate the landmarks’ locations within the VE. We employed a quantifiable measure [15], through the mean cosine of the angle between the vectors representing the landmarks’ positions predicted by the network, namely the centroids of the clusters, and the vectors representing the landmarks’ positions in the virtual world: (x, y) . The average cosine was 0.9996, while 1 would represent a perfect performance.

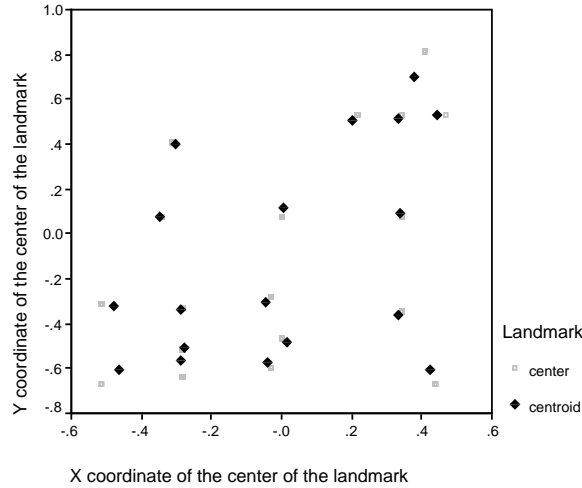


Fig. 6 Original landmarks: centers and the centroids for their predictions

6 Conclusions

This simulation was carried out with the purpose of showing that some abstract aspects related to spatial cognition are learnable. The basic idea is that by mapping an input vector consisting of current Cartesian coordinates together with information about the nearest landmark such as distance to it and the coordinates of its center is sufficient to induce the internal abstractions to predict the next position. Moreover, the network is able to understand the spatial configuration of the VE.

The network predicted correctly the next position together with its nearest landmark at a rate of 67.57%. It was also able to learn the boundaries of the spatial layout, and even to build a cognitive-like map. The spatial representation of the virtual world preserves both topology and metric. The network was also able to assign saliency to landmarks, related to their location e.g. centrality in the room frame, their size and distinctiveness.

A future direction will be to analyze the representations in the hidden layer in order to extract rules or procedural knowledge underlying the navigational behavior. Using neural networks as a tool in studying navigation can be beneficial for user modeling in the area of spatial knowledge acquisition. Permitting a comparative analysis between efficient and inefficient navigational strategies, this methodology could suggest how VEs might be better designed. Based on these results further work will be focused on assisting new users, to improve their spatial abilities in exploring a new VE. After a period of navigation, users could be classified in a cluster according with the navigational patterns [20]. By predicting the user's following trajectory, pertinent advice could be provided to reduce its offset from the desirable "good" trajectory. Thereafter this guidance will improve user exploration. Alternatively, a real-time dynamic reconstruction of the VE could assist the users in their tasks.

References

1. Amant, R.S. and Riedl, M.O. A practical perception substrate for cognitive modeling in HCI. *International Journal of Human Computer Studies*, 55, 1, 2001, 15-39
2. Christos, S. What is a neural network? *Surprise 96 On-Line Journal*, 1, 1996.
3. Darken, R.P. Wayfinding in Large-Scale Virtual Worlds, in *Proceedings of SIGCHI '95*, ACM Press, 45–46.
4. Elman, J.L. Finding structure in time, *Cognitive Science*, 14, 1990, 179–211.
5. Finlay, J. and Beale, R. Pattern recognition and classification in dynamic and static user modelling, in Beale, R. and Finlay, J. (eds.). *Neural Networks and Pattern Recognition in Human Computer Interaction*. Ellis Horwood, 1992.
6. Gabbard, J. and Hix, D. *Taxonomy of Usability Characteristics in Virtual Environments*, Final Report to the Office of Naval Research, 1997.
7. Golledge, G.R. Environmental cognition, in Stokols, D. and Altman, I. (eds.) *Handbook of Environmental Psychology*, vol. 1. John Wiley & Sons, Inc. New York, 1987, 131–175.
8. Golledge, G.R. Human cognitive maps and wayfinding, in Golledge, G.R. (ed.) *Wayfinding behaviour*. John Hopkins University Press, Baltimore, 1-45.
9. Hagner, D.G., Watta, P.B. and Hassoun, M.H. Comparison of Recurrent Neural Networks for Trajectory Generation, In *Recurrent Neural Networks: Design and Applications*, L. Medsker and L. C. Jain (eds.), CRC Press, 1999.
10. Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, 1994.
11. Jacobs, W.J., Thomas, K.G.F., Laurence, H.E., and Nadel, L. Place learning in virtual space II: Topological relations as one dimension of stimulus control. *Learning and Motivation*, 29, 1998, 288-308.
12. Kuipers, B. Modeling spatial knowledge. *Cognitive Science*, 2, 1978, 129–153.
13. Lint, H. van, S.P. Hoogendoorn, H.J. van Zuylen. Freeway Travel Time Prediction with State-Space Neural Networks, Preprint 02-2797 of the 81st Annual Meeting of the Transportation Research Board, Washington D.C., 2002.
14. Lynch, K. *Image of the city*. The MIT Press, Cambridge MA, 1960.
15. Morris, W.C., Cottrell, G.W., Elman, J. A connectionist simulation of the empirical acquisition of grammatical relations, In Wermter, S. and Sun, R. (eds.) *Hybrid Neural Symbolic Integration*. Springer Verlag. 2000.
16. O'Hare, G.M.P., Sewell, K., Murphy, A.J. and Delahunty, T. ECHOES: An Immersive Training Experience. *Proceedings of Adaptive Hypermedia and Adaptive Web-based Systems*. 2000, 179–188
17. Pearlmutter, B. Learning state space trajectories in recurrent neural networks. *Neural Computation*, 1(2): 263-269, 1989.
18. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. *Human-Computer Interaction*. Wokingham, UK: Addison-Wesley. 1994.
19. Psarrout, A, Gong, S., Buxton, H. Modeling spatio-temporal trajectories and face signatures on partially recurrent neural networks. In Proc. *IEEE International Conference on Neural Networks*, pp. 2226-2232, Perth, Australia, Nov 1995.
20. Sas, C., O'Hare, G.M.P., Reilly, R. A Connectionist Approach to Modeling Navigation: Trajectory Self Organization and Prediction. *Proceedings of 7th ERCIM Workshop, User Interfaces for All*. Carbonell, N. and Stephanidis, C. (eds.), 2000, 111-116.
21. Sayers, H.M., Wilson, S., Myles, W., McNeill, M.D.J. Usable Interfaces for Virtual Environment Applications on Non-Immerse Systems, In *Proceedings Eurographics UK*, 2000, 143-150.
22. Sundareshan, M.K., Wong, Y.C. and Condarcur, T. Training algorithms for recurrent neural nets that eliminate the need for computation of error gradients with application to trajectory production problem, In *Recurrent Neural Networks: Design and Applications*, L. Medsker and L. Jain (eds.), CRC Press, 1999.