Identifying User Destinations in Virtual Worlds

Fahad Shah and **Philip Bell** and **Gita Sukthankar**
School of EECS
University of Central Florida
4000 Central Florida Blvd.
Orlando, Florida 32816
sfahad@cs.ucf.edu and phbell@earthlink.net and gitars@eecs.ucf.edu

Abstract

This paper focuses on the identification of human activity patterns in SecondLife (SL), a user-constructed virtual environment. SecondLife allows the users to create a virtual avatar, explore areas constructed by other users, socialize, and conduct financial transactions just as one would in the real world. However unlike the real world, new attractions can be constructed within hours and previous ones often fall into disuse rapidly. Without current information about the state of regions in the virtual world, it is difficult to infer the purpose of the user's actions from location information. In this paper, we present an approach for gathering data on users' activities and building a map of SecondLife annotated with information about activities that the users were able to perform in each region. Using this map, a recommender agent built into the user's heads-up display can present suggestions of other areas to visit based on data collected from previous users. We discuss the the use of five supervised classifiers and report classification results for the map construction portion of the agent.

Introduction

Second Life is a user-constructed virtual environment that allows the users to construct their own 3D world. The game supports a number of personal modes of travel (walking, flying, teleporting) in addition to allowing users to create their own vehicles. Second Life is laid out in a 2D space, SLGrid, composed of regions with each region being hosted on its own server and offers a fully featured 3D environment that is built and shaped by the users who own the land. With the increasing number of attractions, it is difficult for a user to explore all the possibilities and places to visit. This motivates the need for a recommendation system that can suggest places to visit, personalized with the user's activity preferences.

System

To create an activity-annotated map of regions in Second Life, we built a tracker object using the Linden Scripting Language (LSL). Users carrying the object are periodically prompted for information to describe their current location.

Figure 1: Screenshot of a user walking around in Second Life with the tracker object worn as a belt attached to the HUD (displayed as blue belt at the bottom left). The menu displayed at the top-right corner prompts the user to specify a category for the place being visited.

This map is designed to be used by a recommender agent that suggests other destinations, allowing the user to quickly identify regions of interest and minimize wasted navigation time.

The tracker object is a belt that can be worn on the right or left of the avatar that monitors the user's current x,y,z location as shown in Figure 1. The tracker commences operation when the user clicks on the object and periodically prompts the user for an activity annotation to describe the current location (Educational, Residential, Recreational, Shopping, Camping, Other). After a certain time limit, all of the information collected is emailed to an external account.

The tracker's emails are processed by a separate Java application that downloads the email from the server, and stores the information in a MySQL database. Once we have the data from the individual user, we consolidate the data collected from all the user logs, incrementally, across all the users in a single un-normalized table for classification.

We use the Waikato Environment for Knowledge Analysis (Weka)[1] to classify the data. Using Weka we explored the performance of five supervised learning algorithms to pre-

[1]Weka home page: http://www.cs.waikato.ac.nz/ml/weka/

Table 1: Classification Performance

Classification Scheme		Accuracy	Training Time (s)	Statistical Ranking Wins− Losses	Wins	Losses
Bayesian schemes	Bayes Net	94.72%	0.06 s	2	2	0
	Naive Bayes	84.24%	0 s	-4	0	4
Trees	C4.5	92.81%	0.01 s	-1	1	2
Functional schemes	MLP	93.48%	10.96 s	0	1	1
Lazy schemes	**IBK**	**95.00%**	**0 s**	**3**	**3**	**0**

dict the user's current activity at a given location using the activity-annotations collected with the tracker. The user's future destination is then predicted using three independent models (built separately from the data in the database), each employing the M5P (Quinlan 1992) algorithm for numeric prediction. When predicting the values, the time information is assumed to be one minute after the last timing information logged from the user track in the database while all the other information is taken to be the same as that for the last user record. With this information, we can make a prediction for the category using one of the supervised category prediction classifiers and return a suggestion to the user about where to visit based on the appropriate category information.

Results

Each model was trained using the data collected by the tracker augmented with the category labels interactively entered by users. The goals of the task were to: (1) predict the destination (x, y, z) of the user; and (2) predict the category for a given location in the virtual world.

For category prediction, we evaluated the classification performance and learning time of the following five algorithms: C4.5 Decision Tree (Quinlan 1993), Bayesian belief network (Bouckaert 1995), Naive Bayes (John & Langley 1995), Multilayer Perceptron (Mitchell 1997) and IBK (k-nearest neighbor) (Aha, Kibler, & Albert 1991). Each model is trained using the data collected by the tracker augmented with the category labels interactively entered by users. The experiments reported here employed a data set collected during a pilot study of five users. Classification accuracy was estimated using 10-fold cross-validation, and statistical analysis performed using corrected t-test (adjusted for folds). The win-loss ranking for each algorithm counted the number of times that a given classifier was significantly better than others (using significance of 0.05). The time required to train each classifier was also recorded. Table 1 summarizes these results for each of the five classifiers.

Since there are six category labels, the baseline (chance) performance on this task is 16.7%. All of the algorithms perform much better than this baseline, confirming our belief that there is significant spatial structure for user-created attractions in Second Life. The worst-performing algorithm, Naive Bayes, assumes that the features are conditionally independent, which is clearly not the case. MLP does quite well, and is robust to noise in the input data (a reason for why MLPs are frequently used in sensor applications). On our task, MLP is not the best performer, and in light of its longer training time, there appears to be little merit in using it here. The Decision Tree attempts to minimize the entropy over training examples and does post pruning based on the error estimate. While its classification accuracy on our task is high, the Decision Tree is statistically outranked by several of the other methods. The Bayes Net is one of the top two performers and offers the benefit of enabling us to incorporate prior knowledge about Second Life's spatial layout in a principled manner. The best performance in our study was achieved by the k-nearest neighbor (IBK) algorithm, which is a lazy learner that simply stores training instances and generates a classification label at query time using a non-parametric density estimate of labels in the local neighborhood. The algorithm is very robust to noise, both in features and in class labels, and is amenable to incremental updates.

Conclusion

This paper presents a system for mapping virtual worlds using activity-based annotations provided by users through a customized tracking object. The user's destination (x,y,z) is predicted in real-time using three independent models, each employing the M5P algorithm for numeric prediction. We explore the use of five supervised classifiers, Bayes belief networks, Naive Bayes, Multilayer Perceptron (MLP), K-nearest neighbor and decision trees, to construct an activity-annotated map of the Second Life from user data. A ranking of the classification algorithms based on the statistical differences in classification show that IBk performs the best, followed by Bayesian belief networks.

These predictions of the user's activity patterns could be used as a basis for creating per-user experiences and user-targeted advertising. Many people use the internet to create a social presence, through blogs, avatars, and social networking sites. This presents an opportunity for researchers to collect rich user data from these interactions to create a more personalized and user-friendly experience. We believe that combining the information that we get from the user's SL experience with the information measures from the user's interests in real life is a promising area for future research.

References

Aha, D. W.; Kibler, D.; and Albert, M. K. 1991. Instance-based learning algorithms. *Machine Learning* 1.

Bouckaert, R. 1995. *Bayesian Belief Networks: from Construction to Inference*. Ph.D. Dissertation, University of Utrecht.

John, G. H., and Langley, P. 1995. Estimating continuous distributions in Bayesian classifiers. In *Proceedings of Conference on Uncertainty in Artificial Intelligence*.

Mitchell, T. M. 1997. *Machine Learning*. McGraw-Hill Science/Engineering/Math. chapter 4, pp.95–103.

Quinlan, J. R. 1992. Learning with continuous classes. 343–348. World Scientific.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.