Comparing Close Destination and Route-Based Similarity Metrics for the Analysis of Map User Trajectories

Ali Tahir[1], Gavin McArdle[2], and Michela Bertolotto[1]

[1] School of Computer Science and Informatics, University College Dublin (UCD)
Dublin, Ireland
(ali.tahir, michela.bertolotto)@ucd.ie

[2] National Centre for Geocomputation, National University of Ireland Maynooth (NUIM)
Maynooth, Ireland
gavin.mcardle@nuim.ie

Abstract. Movement is a ubiquitous phenomenon in the physical and virtual world. Analysing movement can reveal interesting trends and patterns. In the Human-Computer Interaction (HCI) domain, eye and mouse movements reveal the interests and intentions of users. By identifying common HCI patterns in the spatial domain, profiles containing the spatial interests of users can be generated. These profiles can be used to address the spatial information overload problem through map personalisation. This paper presents the analysis and findings of a case study of users performing spatial tasks on a campus map. Mouse movement was recorded and analysed as users performed specific spatial tasks. The tasks correspond to the mouse trajectories produced while interacting with the Web map. When multiple users conduct similar and dissimilar spatial tasks, it becomes interesting to observe the behaviour patterns of these users. Clustering and geovisual analysis help to understand large movement datasets such as mouse movements. The knowledge gained through this analysis can be used to strengthen map personalisation techniques. In this work, we apply OPTICS clustering algorithm to a set of map user trajectories. We focus on two similarity measures and compare the results obtained with both when applied to particular saptial tasks carried out by multiple users. In particular, we show how route-based similarity, an advanced distance measure, performs better for spatial tasks involving scanning of the map area.

1 Introduction

Recently there have been huge advances in spatial technologies. The use of Web maps and positioning technologies are now ubiquitous. The drive to generate spatial content has led to a major information overload problem [1], in which finding timely and relevant spatial information becomes a challenge. This makes it difficult for the users to choose and filter content to match their current needs.

S. Liang, X. Wang, and C. Claramunt (Eds.): W2GIS 2013, LNCS 7820, pp. 117–128, 2013.

Therefore, there is a need to adapt Web map contents into personalised maps by understanding the user interests. Analysing spatial interaction patterns offers an approach to resolve this.

The research presented in this paper focuses on data generated by HCI in a Web map environment, in particular, the paths generated by users through computer mouse movements. These paths, called trajectories, represent a series of mouse cursors locations. Mouse trajectories reflect usage patterns and activities, which help to predict future movements and interests. In order to analyse our trajectory data, we use Visual Analytics (VA) [2], which promotes cognition and knowledge discovery through data visualisation in a geospatial environment combined with analysis tools such as machine learning, data mining and clustering.

In our approach, clustering of mouse trajectories (obtained as 2D mouse movements) is an integral component in order to determine the usage patterns. We apply the OPTICS clustering algorithm [3], a density-based clustering approach. Such an algorithm requires a metric upon which to determine similarity between trajectories. We compare two similarity metrics, namely close destination and route-based [4], and assess when it is advantageous to use them.

In our previous work [5] we performed a preliminary evaluation based on a small experiment (involving 12 users) using the close destination measure. In this paper we present a larger and more comprehensive user trial in which 27 users participated. Each user was asked to perform ten spatial tasks. The total trajectories collected from all users were analysed in order to see common patterns with the help of visual analysis and clustering techniques. As we are interested in identifying types of user, we analysed multiple users over individual tasks using both route-based and close destination simialrity measures to identify common patterns. Maps of tasks and users are generated to validate the clustering results. The results show that clustering and geovisual analysis is an effective technique to understand user patterns and behaviours. They also show how route-based measures perform better for certain types of tasks.

The goal of this paper is to demonstrate the use of a geovisual analysis tool to analyse user mouse trajectories. In particular, the work assesses the benefits of different distance measures for clustering such trajectories. The results highlight how different users approach particular tasks and will provide input into map personalisation algorithms to resolve the spatial information overload problem.

The remainder of this article is organised as follows. To provide the background for our work, some related literature is presented in Section 2. The overall approach which includes a description of the clustering algorithm, distance measures and spatial tasks is described in Section 3. Section 4 presents experimental results and the evaluation we have conducted. Finally, Section 5 outlines the conclusions and directions for future work.

2 Related Work

Activity recognition from movement data (in particular, spatial trajectories) is challenging due to the huge amount of datasets available. Zhu et al. [6] define this activity recognition as "the process to extract high-level activity and goal related information from low-level sensor readings through machine learning and data mining techniques". The authors represent trajectory-based activity recognition into three levels. The first level inputs the location from a sensor (for example, GPS, WiFi, cameras). At the next level, an activity is recognised, such as transportation modes. As an example of the final (highest) level, Microsoft GeoLife [1] a Location-based Social Network (LBSN) application, has a transportation detection system which categorises a GPS trajectory into various activities (such as walking, biking, driving and onBus). These transportation modes facilitate activity recognition. GeoLife aims to provide social connectivity between people using their trajectories.

The advances in location-based positioning technologies such as GPS as well as the growth of mobile and wireless technologies have enabled movement data,in the form of spatial trajectories, to be collected. These trajectories can be collected in both outdoor and indoor [7] environments and correspond to free [8, 9] and constrained movements [10, 11]. For example, a road network corresponds to a constrained movement, while the movement of an animal can be termed as free movement. Zheng and Zhou [12] provide several movement data examples. They categorise movement of people into active and passive recordings. For example, active recording occurs when travellers share their travel routes with their friends to strengthen social connectivity which is evident in the context of Location-based Social Networks (LBSN) applications. On the other hand, a user carrying a mobile phone unintentionally generates spatial trajectories. These trajectories correspond to a sequence of cell tower IDs. Other categories include the mobility of vehicles (which can be used for traffic analysis and resource allocation) and the mobility of animals and natural phenomenons (such as migratory birds research, hurricanes and tornadoes).

While most researchers focus on physical trajectories, mouse, eye and touch gestures on a computer screen are also forms of movement which generate specific trajectories. When interacting with a spatial application, such as a Web map, these trajectories can be analysed to examine Human-Computer Interaction (HCI), the study of interaction between humans and machines [13]. We are interested in analysing trajectories generated by HCI in our research, particularly in the spatial interaction domain.

Typical activity recognition involves monitoring a single as well as multiple user activities. Zhu et al. [6] present a detailed account of single and multiple users activity recognition from trajectory data. The study shows that much work has been done on single users while less attention has been given to multiple user activity analysis. Single user activity recognition corresponds to the analysis of an individual's user history in order to predict future trends. In general, single user activity can be recognised through supervised methods, unsupervised methods and frequent pattern mining approach [6]. Supervised learning methods typically input trajectory data as well as activity labels. The intention is to use a suitable classification model such as Bayesian networks, Hidden Markov Model

[1] http://research.microsoft.com/en-us/projects/geolife/

(HMM) and decision trees, in order to predict the activity of the trajectory in question. These methods attempt to identify activities within a trajectory, for example, significant places, stop rate, velocity change, etc. Unsupervised methods on the other hand do not take activity labels into account, but instead discover patterns directly from the trajectory data [6], for instance, by applying clustering methods. For this reason, in the case of mouse trajectories, unsupervised methods are found to be more relevant. To the best of our knowledge, mouse trajectories on a map interface have not been studied in order to understand user's intentions and behaviours.

In our previous work, we have applied spatial clustering to mouse trajectories in order to identify usage patterns [5]. The approach successfully identified spatial tasks with the help of clustering while some outliers were also detected. As an extension to spatial clustering, we presented clustering based on temporal information which considers speed and acceleration at each location in a trajectory to describe behaviour [14]. A detailed discussion on clustering algorithms suitable for trajectories as well several distance measures is presented in [5, 14]. The research presented here is an extension to this. We adapt an advanced similarity measure (route similarity) [4] and use this as a spatial distance measure in OPTICS clustering. In previous work, we used close destination distance measures to find the similarity between trajectories. This function considers the end points of each trajectory while computing the trajectory similarity. Now we compare the distance between each point of each trajectories in order to produce a new overall similarity score.

3 Approach

In this section we present our approach. Firstly, we conducted a user trial in which users performed several spatial tasks. The trials were conducted in an unsupervised manner, similar to a previous set of trials described in [5]. The participants had to register in order to start the spatial tasks. The Web interface, designed for this purpose, was deemed to be as user friendly as possible as shown in Figure 1. The interface contains different components on a Web page including the spatial task description, an input area for answers, a mapping component and a map legend. The geographic area selected was that of the University College Dublin (UCD) Belfield Campus.

The 10 spatial tasks presented to users are listed below, while the Web interface is shown in Figure 1.

- Task 1: As a new student, you are required to open up a student account with the bank located at the campus. Find the bank on the campus map. On your way back, you would also like to get some food. Which is the nearest restaurant to the bank?
- Task 2: You finish a lecture at the John Henry Newman Building (Arts) and need to catch a bus to the city centre (all bus stops connect to the city). Find the bus stop most convenient to you on campus. How many bus stops are there on campus?

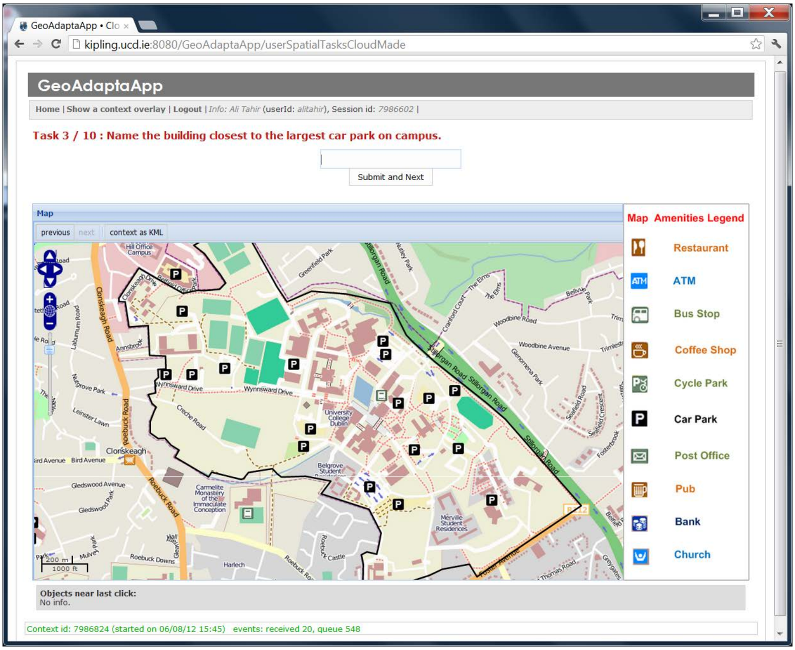

Fig. 1. Web interface for experiments

- Task 3: Name the building closest to the largest car park on campus.
- Task 4: In order to attend a service at the church located on campus, you need to find the closest car park to the church. Find your way to this car park if you are driving from the Stillorgan Road/N11 entrance.
- Task 5: You need to attend a graduation ceremony at UCD. You will be driving to UCD and enter from the Stillorgan Road/N11 entrance. Locate the reception and O'Reilly Hall and find the shortest route between the two.
- Task 6: You need to drive to the student bar in the evening. You are required to enter the UCD campus from the Wynsward Drive entrance. Follow the path to find the closest car park to the student bar. How many pubs are shown on the map?
- Task 7: You finish a lecture in Computer Science and Informatics (CSI) building. You need to meet your friend in front of the Health Science Centre and go together to the James Joyce Library (close to the central largest lake - coloured in blue) in order to return a book. Plan the route.
- Task 8: In order to post a letter to your friend, plan the route to cycle from Charles Institute located at North West of the central lake (coloured in blue) to Belfield Post Office. Find the nearest bicycle parking stand to the post office.

– Task 9: You need to meet your friends at the sports centre building (coloured in green). How would you get to the Glenomena student residence (south east of the central lake) from the sports centre?
– Task 10: Count the number of roads crossing the UCD boundary (as outlined by the black line).

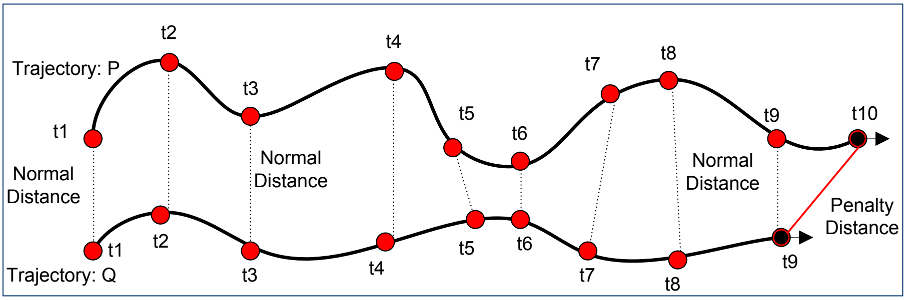

Fig. 2. Normal distance and penalty distance between two trajectories. In case of non-corresponding points between two trajectories, penalty distance is calculated (shown in red), and added to the overall distance between two trajectories.

The trajectories produced by the user mouse movements were analysed using the OPTICS clustering algorithm. OPTICS [3] produces an ordering of a dataset while it searches for a core distance and a reachability distance of each trajectory with respect to its predecessor. OPTICS outputs a reachability plot. From the output plot, groupings can be obtained by choosing an appropriate threshold value of reachability distance. Let ρ = object from a dataset D, ε = distance threshold, $N\varepsilon$ (ρ) = ε-neighborhood of object ρ, *minPts* = natural number, *minPts-distance*(ρ) = distance from ρ to its *minPts* neighbor. The core distance (CD) is defined as:

$$CD = \begin{cases} Undefined, & \text{if } Card(N\varepsilon(\rho)) < minPts \\ minPts\text{-}distance(\rho), & \text{otherwise} \end{cases}$$

The *core distance* is the smallest distance ε between ρ and an object in its ε-neighborhood such that ρ would be a core object. The *core distance* is *Undefined*, otherwise. For reachability distance, let ρ and o = objects from a dataset D, $N\varepsilon$ (o) = ε-neighborhood of object o, *minPts* = natural number. The reachability distance (RD) of ρ with respect to o is defined as:

$$RD = \begin{cases} \mathit{Undefined}, & \text{if } Card(N\varepsilon(o)) < minPts \\ max(core\text{-}distance(o), distance(o, \rho)), & \text{otherwise} \end{cases}$$

Thus, the reachability distance of ρ is the smallest distance such that ρ is directly density-reachable from a core object o. Otherwise, if o is not a core object, even at the generating distance ε, the reachability distance of ρ with respect to o is *Undefined*.

In order to apply this clustering algorithm, an appropriate similarity measure is required. The similarity measure is used with the clustering algorithm along with a distance threshold and the minimum number of neighbours required to form a cluster. Route similarity [4] is a complex (computationally expensive) distance measure that computes the geographical distance between two trajectories. Moreover, this distance function deals with incomplete trajectories and with more significant positioning errors. Although mouse trajectories do not have positioning errors, they are uncertain and hold frequent sequences of movements. This function is also more suited to unequal time intervals between records (the case of mouse trajectories). The function repeatedly searches for the closest pair of positions in two trajectories. It computes two derivative distances: mean distance between the corresponding positions and a penalty distance for the unmatched positions. The penalty distance is increased if a position is skipped and decreased if the corresponding position is found. The final distance is the sum of two derivative distances. We have opted to use this measure in order to compute the distance between two trajectories. Figure 2 illustrates the concept of normal distance and the penalty distance between two trajectories. The next section presents results and evaluation we have conducted.

4 Evaluation

In order to demonstrate the effectiveness of route similarity with OPTICS, we carried out a series of analysis tasks. Although our main interest lies with route similarity applied to individual tasks, initially we show the effects of applying close destination clustering to all trajectories produced from the user trial as the results provide an example of a situation where this is not an effective metric. Additionally, the approach highlights how to use and interpret the output from the OPTICS algorithm which we also used to compare the route-based and close destination clustering of trajectories from specific tasks.

Figure 3 shows the results of applying OPTICS clustering with close destination similarity measure to all meaningful trajectories collected during the user trial. While a 100 percent completion rate by all participants would yield *270* trajectories, not all users interacted with the map during trials, or completed all tasks and so *258* trajectories were recorded. Close destination clustering requires a distance parameter and a minimum number of neighbours to be specified, these were set to 1000m and 7 neighbours respectfully. The graph produced reveals the clusters present in the trajectories. Clusters can be identified using the rules provided by the authors of [2]: "the first point of a cluster (called the start of the cluster) is the last point with a high reachability value, and the last point of a cluster (called the end of the cluster) is the last point with a low reachability value". Using this rule a cluster starting at position *21* and ending at position *40* can be extracted. Similarly, another cluster starting at position *97* and ending at position *113* can be identified.

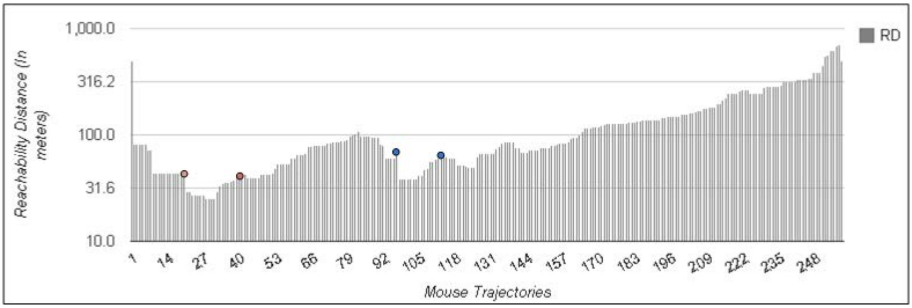

Fig. 3. Similarity measure: close destination, distance: 1000 meters, neighbours: 7, Trajectories: 258

Once extracted, the trajectories can be visually analysed, for example, in Figure 4 all trajectories forming a cluster are assigned the same colour. The black symbol indicates the end of each trajectory. Overall, this clustering approach did not produce clusters corresponding to specific tasks or specific users. In this approach trajectories which terminate in close proximity to each other are deemed similar and so should cluster trajectories based on the task they are part of (i.e. terminating at the specific spatial feature of interest in a given task). However, due to the small study area, the trajectories terminate in close proximity for many tasks and so few distinct clusters were identified. Importantly, several of the tasks do not require the identification of a specific spatial feature but require the user to scan the map, count features or identify routes. As a result, similarity based on close destination is not always effective and is task specific.

As our interest is in individual tasks, we applied route-based similarity measures, with OPTICS clustering on a task-by-task basis and compared it to the results obtained using close destination. Initially, we considered task 1 (described in Section 3). This task required users to locate a bank and a nearby restaurant on the campus map. Task 1 was performed by *25* users which included *15* familiar and *10* non-familiar (with the map area) users. Figure 5 (a) shows the outcome when OPTICS was applied to the *25* user trajectories using a close destination similarity measure along with the chosen input parameters (distance: 300m, neighbours: 5). It can be seen that the reachability chart in this case does not produce a definitive set of clusters. This is due to the nature of the task involving scanning the map and identifying *2* locations. Figure 5 (b) shows the results using a route similarity measure (distance parameter: 2500m, neighbours: 5). Here, three distinct clusters are identified. These clusters are shown as an overlay on the study area in Figure 6. The clusters clearly identify the approach users took to complete the tasks and can be broadly broken down into users who were familiar and unfamiliar with the campus.

To further demonstrate the effectiveness of using a route similarity metric we also analyse the results for task 10. In task 10, users were required to count the number of roads entering the UCD campus. This is an example of a situation in which close destination similarity is not effective because users are free to

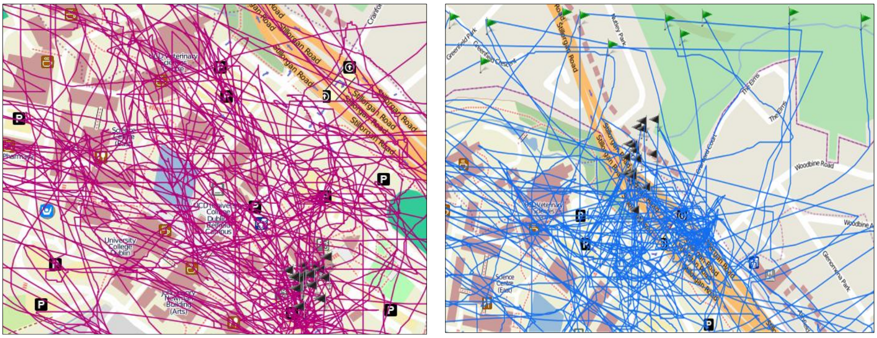

Fig. 4. (a) Cluster showing 22 trajectories starting from location 21 and ending at 40. (b) Cluster showing 22 trajectories starting from location 97 and ending at 113.

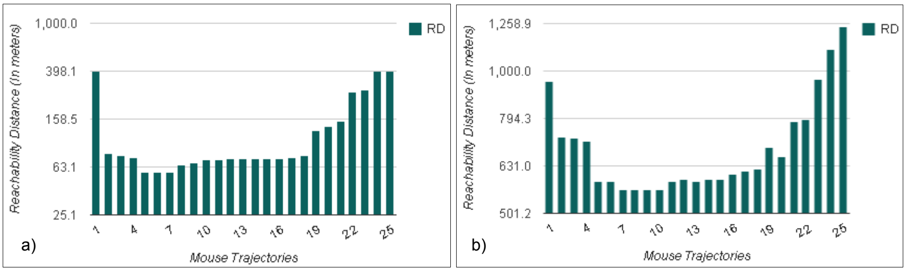

Fig. 5. Task 1- a) Similarity measure: close destination, distance: 300 meters, neighbors: 5, trajectories: all b) Similarity measure: route similarity, distance: 2500 meters, neighbors: 5, Trajectories: all

start and end their trajectories at arbitrary locations. This task involves scanning the map to identify roads crossing the campus boundary. This task was performed by all *27* users. Both close destination and route similarity measures were used in conjunction with OPTICS and the results are shown in Figure 7. The parameters used were a distance of 2500m and 5 neighbours. The close destination metric did not identify a distinct set of clusters whereas route similarity clearly identified two predominant groups of trajectories. These two clusters can be visually analysed in Figure 8. The first cluster (Figure 8 (a)) is found to be very neat where users precisely followed the path along the boundary line of the campus map. These trajectories correspond to users who were familiar with the campus and are in contrast to the trajectories in Figure 8 (b) where there is seemingly random mouse movement and trajectory shape. These represent users who indicated they were unfamiliar with the campus.

In conclusion, the use of a close destination similarity metric is not effective for all tasks. This is especially true for tasks which do not have a specific goal or target and tasks taking place in a small geographic area. For scanning tasks,

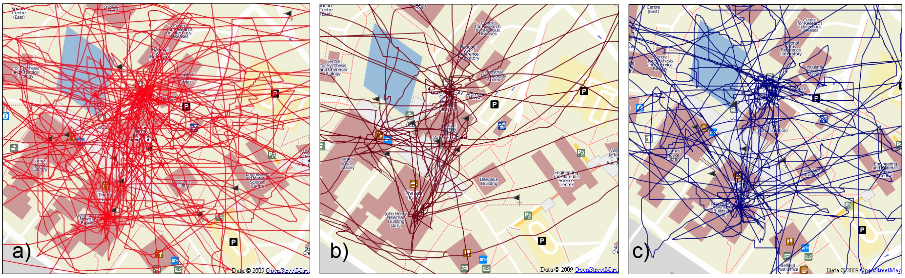

Fig. 6. Task 1- a) First cluster (1-11) b) Second cluster (12-18) c) Third cluster (19-25)

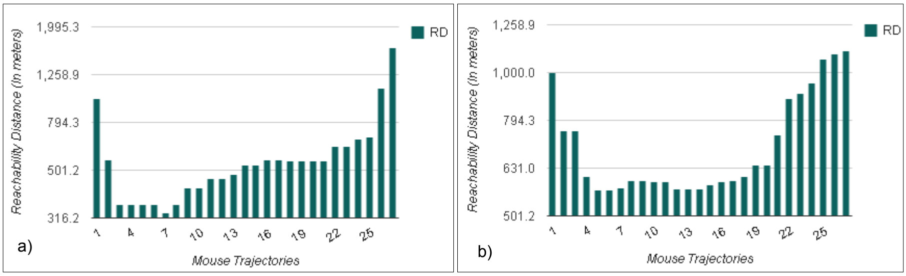

Fig. 7. Task 10- a) Similarity measure: close destination, distance: 2500 meters, neighbors: 5, trajectories: all b) Similarity measure: route similarity, distance: 2500 meters, neighbors: 5, trajectories: all

it is more advantageous to use a similarity metric which looks at the complete path of a trajectory such as route-based similarity. The results here show that this is particularly useful for examining the behaviour of users and determining their familiarity with the study area and the task at hand.

5 Conclusion and Future Work

This paper presents clustering and geovisual analysis of mouse trajectories collected from users performing spatial tasks. By analysing users as they perform such tasks, parameters which can be used in map personalisation can be collected. In this paper we describe a tool we have developed which applies OPTICS clustering and geovisualisation to analyse HCI data. The paper focuses on the use of similarity measures for comparing and clustering trajectories of users completing specific tasks. In particular we examine close destination and route-based similarity measures. Close destination similarity identifies trajectories which terminate in the same region whereas route similarity measures consider the complete trajectory path. We apply these techniques to *258* mouse trajectories collected during a user trial in which participants had to complete *10*

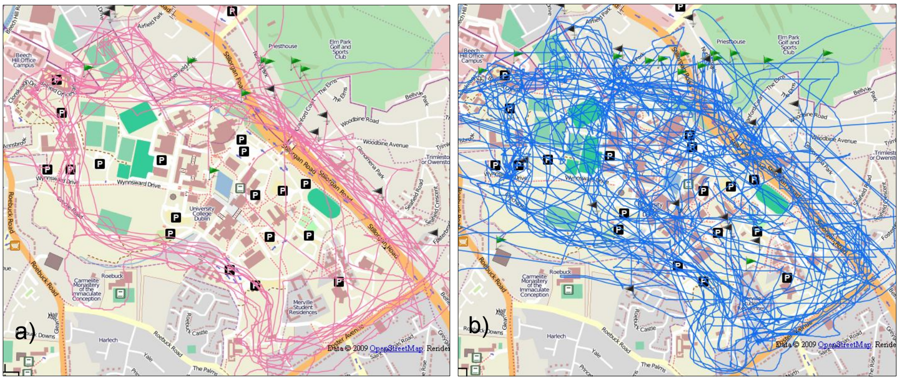

Fig. 8. Task 10- a) Small cluster (1-10) b) Large cluster (11-27)

specific spatial tasks using a Web map. The results highlight the benefits of using the route-based approach for tasks which involve scanning the map or identifying more than one spatial feature, and show it is effective at highlighting users' familiarity with the study area. As route-based clustering considers the trajectory shape, it also shows the approach users take to complete a task. This was validated through visual analysis of cluster assignment and via a questionnaire presented to users.

The results reinforce the importance of choosing the correct clustering approach based on the task at hand. Individual user trajectories for several spatial tasks can be clustered to reveal usage patterns for specific users or participants. The participants can also be distinguished based on their map usage experience and the area familiarity. Similarly mouse trajectories can be clustered by taking map scale factor into account which at the moment is part of the visual analysis. To further strengthen and validate the clustering techniques, heat maps and speed maps can be combined with the results [15]. Overall, the paper has demonstrated an approach for the analysis of map based HCI. The next stage is to apply the results to map personalisation. This can be achieved by adapting map content based on the users' interests, ability and approach as determined through the cluster analysis presented here.

Acknowledgements. Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge this support.

References

[1] Yang, Y., Claramunt, C.: A Hybrid Approach for Spatial Web Personalization. In: Li, K.-J., Vangenot, C. (eds.) W2GIS 2005. LNCS, vol. 3833, pp. 206–221. Springer, Heidelberg (2005)

[2] Thomas, J., Cook, K.: A visual analytics agenda. IEEE Computer Graphics and Applications 26, 10–13 (2006)
[3] Ankerst, M., Breunig, M.M., Kriegel, H.P., Sander, J.: OPTICS: Ordering Points to Identify the Clustering Structure. SIGMOD Rec. 28, 49–60 (1999)
[4] Andrienko, G., Andrienko, N., Wrobel, S.: Visual analytics tools for analysis of movement data. SIGKDD Explorations Newsletter - Special Issue on Visual Analytics 9, 38–46 (2007)
[5] Tahir, A., McArdle, G., Bertolotto, M.: Identifying specific spatial tasks through clustering and geovisual analysis. In: 2012 20th International Conference on Geoinformatics (GEOINFORMATICS), pp. 1–6. IEEE (2012)
[6] Zhu, Y., Zheng, V.W., Yang, Q.: Activity recognition from trajectory data. In: Zheng, Y., Zhou, X. (eds.) Computing with Spatial Trajectories, p. 180. Springer, New York (2011)
[7] Jensen, C.S., Lu, H., Yang, B.: Indexing the Trajectories of Moving Objects in Symbolic Indoor Space. In: Mamoulis, N., Seidl, T., Pedersen, T.B., Torp, K., Assent, I. (eds.) SSTD 2009. LNCS, vol. 5644, pp. 208–227. Springer, Heidelberg (2009)
[8] Güting, R., Böhlen, M., Erwig, M., Jensen, C., Lorentzos, N., Schneider, M., Vazirgiannis, M.: A foundation for representing and querying moving objects. ACM Transactions on Database Systems (TODS) 25, 1–42 (2000)
[9] Forlizzi, L., Güting, R., Nardelli, E., Schneider, M.: A data model and data structures for moving objects databases, vol. 29. ACM (2000)
[10] Güting, H., de Almeida, T., Ding, Z.: Modeling and querying moving objects in networks. The VLDB Journal, The International Journal on Very Large Data Bases 15, 165–190 (2006)
[11] Speičvcys, L., Jensen, C., Kligys, A.: Computational data modeling for network-constrained moving objects. In: Proceedings of the 11th ACM International Symposium on Advances in Geographic Information Systems, pp. 118–125. ACM (2003)
[12] Zheng, Y., Zhou, X.: Computing with Spatial Trajectories. Springer-Verlag New York Inc. (2011)
[13] Dix, A., Finlay, J., Abowd, G.: Human-Computer Interaction. Prentice Hall (2004)
[14] McArdle, G., Tahir, A., Bertolotto, M.: Spatio-temporal clustering of movement data: An application to trajectories generated by human-computer interaction. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences I-2, 147–152 (2012)
[15] Tahir, A., McArdle, G., Bertolotto, M.: A geovisual analytics approach for mouse movement analysis. International Journal of Data Mining, Modeling and Management. Special Issue on Spatial Information Mining, Modeling and Management (2012) (in press)