Trajectory Similarity of Moving Objects[1]

Marc van Kreveld, Jun Luo

Department of Information and Computing Sciences, Utrecht University
{marc, ljroger}@cs.uu.nl

Abstract. Computing the similarity between two trajectories is the key to similarity based clustering. Most research concentrates on shape similarity, and speed does not play a role. Furthermore, only the similarity between whole trajectories of two moving objects is considered. In this paper we use a new definition of the (dis)similarity between two trajectories which considers both temporal effects and (dis)similarity of subparts of trajectories. We present four variants of this problem and provide algorithmic solutions.

1 Introduction

With the widespread use of location-based devices like GPS receivers, trajectories of moving objects are captured every day and a huge amount of trajectory data is produced. This opens up new possibilities for the analysis of trajectories of moving objects like people, animals, vehicles, and hurricanes (Laube 2005). One major method of spatio-temporal data mining is cluster analysis. The foundation of similarity-based clustering is to compute the similarity between two trajectories. There are many methods for this (Fu et al. 2005, Vlachos et al. 2002a, Vlachos et al. 2002b, Yanagiswa et al. 2003). However, most methods do not consider temporal effects like speed of the moving object. In many situations, speed plays an important role. For example, figure 1 shows two hurricane trajectories τ_1 and τ_2 that have the same shape and a similar location. But the development of the speed of the two trajectories is very different: hurricane 1 is fast at first, and then moves slowly, whereas this is the other way around for hurricane 2. We need a similarity measure to reflect the difference of these two trajectories. In this paper, we use a definition of dissimilarity between two trajectories that can capture this. Furthermore, we show how to compute the most similar subtrajectories of two trajectories according to this definition. The most similar subtrajectories are useful for clustering where deviating behavior at the ends should be ignored, and for prediction of the continuation of a trajectory.

[1] This research has been partially funded by the Netherlands Organisation for Scientific Research (NWO) under FOCUS/BRICKS grant number 642.065.503

Florian Probst and Carsten Keßler (Eds.): GI-Days 2007 - Young Researchers Forum.
IfGIprints 30. ISBN: 978-3-936616-48-4

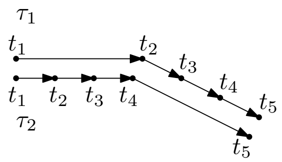

Figure 1: Shape similarity of two hurricane trajectories.

2 PROBLEM DEFINITION AND RESULTS

The trajectory of a moving object is a continuous function $\tau(t)$ of time t such that given a time instant t, it returns the position of the moving object. In reality, the moving object trajectory is recorded by a finite set of observations at discrete times $t_1, t_2, \ldots, t_n$. In this paper we model the trajectory τ of a moving object as a polygonal line with n vertices, and the speed of the object along any line segment is constant.

The dissimilarity, or distance, between the part of τ_1 starting at time t_s and the part of τ_2 starting at time $t_s + t_{shift}$, with duration $T > 0$ is defined as:

$$D(t_s, t_{shift}, T) = \frac{\int_{t_s}^{t_s+T} d(\tau_1(t), \tau_2(t + t_{shift}))dt}{T}$$

where $d(.,.)$ is the Euclidean distance between τ_1 and τ_2 at the times of their arguments. We are interested in subtrajectories with the smallest dissimilarity for given values of T, t, and t_{shift}.

There are four variants of this problem of interest, depending on whether T is fixed and/or whether the starting times of the subtrajectories are same:

- The duration T is fixed and starting times are the same. For example, we want to find the most similar subtrajectory of exactly 3 hours in two animal trajectories recorded during a week.
- The duration T is not fixed but it must be no less than some minimum length T_{min}, and the starting times are the same.
- The duration T is fixed and the starting times may be different. For example, we want to find the most similar subtrajectories of two hurricanes for the duration of exactly four days, but the hurricanes occurred at different times (and may have different lengths).
- The duration T is not fixed but $T \geq T_{min}$, and the starting times may be different.

It is clear that computationally, the first problem will be the easiest and the last one will be the hardest. Figure 2 shows that a time shift can help to find a subtrajectory with smaller distance. If the duration is specified to be two time units, then without a time shift, interval $[t_2, t_4]$ will be most similar, but with a time shift, $[t_7, t_9]$ on τ_1 is more similar to $[t_8, t_{10}]$ on τ_2.

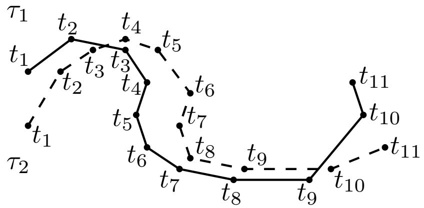

Figure 2: Subtrajectory similarity with time shift.

Figure 3 shows in a one-dimensional example that allowing a longer duration than T_{min} can give a lower value of D, because the dissimilarity measure represents the *average* distance over the time intervals. Assume that no time shift is allowed. If the duration T is fixed to be three units, then $[t_2, t_5]$ gives the lowest D-value, namely 8/3. If the duration may be longer, then $[t_7, t_{11}]$ gives the lower value of 5/2.

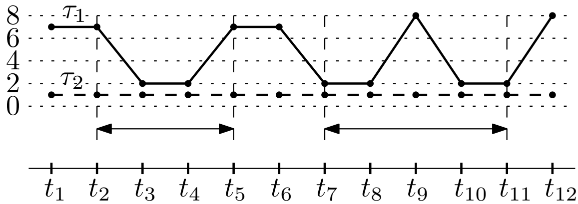

Figure 3: Subtrajectory similarity with non-fixed duration.

We analyze the four versions of the minimum dissimilarity subtrajectory problem and can solve the first version in O(n) time, and the second version in O(n2) time, where n is the number of vertices of the trajectories[1]. It is independent of the specified duration of the subtrajectories. The third and fourth version cannot be solved exactly by analytical means, because the equations that must be solved for minimization are high-degree polynomials. We therefore use approximation, and solve the third and fourth versions in $O(n^2)$ and $O(n^3)$ time, respectively.

[1] The full paper including the details of algorithms and analysis can be downloaded from http://people.cs.uu.nl/ljroger/TrajSim.pdf

REFERENCES

Z. Fu, W. Hu and T. Tan (2005). "Similarity Based Vehicle Trajectory Clustering and Anomaly Detection." In Proc. Int. Conf. on Image Processing (ICIP'05), 2:602-605.

P. Laube (2005). "Analysing Point Motion - Spatio-Temporal Data Mining of Geospatial Lifelines." Ph.D. Dissertation, University of Zurich.

M. Vlachos, G. Kollios and D. Gunopulous (2002a). "Discovering Similar Multidimensional Trajectories." In Proc. 18th Int. Conf. on Data Engineering (ICDE'02), 285-289.

M. Vlachos, D. Gunopulous and G. Kollios (2002b). "Robust Similarity Measures for Mobile Object Trajectories." In Proc. 13th Int. Workshop on Database and Expert Systems Applications (DEXA'02), 721-726.

Y. Yanagisawa, J. Akahani and T. Satoh (2003). "Shape-Based Similarity Query for Trajectory of Mobile Objects." In Proc. 4th Int. Conf. on Mobile Data Management, 63-77.