ACM SIGSPATIAL GIS Cup 2017 - Range Queries under Fréchet Distance

Martin Werner[1], Dev Oliver[2]
[1]German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF),
[2]Environmental Systems Research Institute (Esri)
martin.werner@dlr.de, doliver@esri.com

Abstract

The 25th ACM SIGSPATIAL GIS Conference on Advances in Geographic Information Systems was held in November 2017. In conjunction with the main conference, we organized the 6th GIS-focused algorithm competition called the ACM SIGSPATIAL GIS Cup 2017. The contest was about calculating range queries using the Fréchet distance of trajectories in mobility datasets.

1 Introduction

The ACM SIGSPATIAL GIS Conference addresses all sorts of topics in the fields of geographic information systems since 25 years. In addition to scientific result presentations in the form of papers, short papers, posters, and demos, the conference started to acknowledge the art of algorithm design and implementation very much from a practical perspective through the ACM SIGSPATIAL GIS Cup. Each year, a challenging yet well-researched computational topic is chosen and participants shall solve the given problem in high accuracy, quality, and performance.

For this year's conference, the organizers decided to highlight the Fréchet distance of trajectories. Geospatial trajectories have become an invaluable source of information in spatial analysis, urban computing, transport network research, and map construction. However, comparing trajectories is – in general – quite hard. In this context, a trajectory represents the continuous movement of an object through space. For practical reasons, however, trajectories are usually represented as a sequence of time-stamped spatial locations $T = [(p_1, t_1), (p_2, t_2), \ldots, (p_m, t_m)]$ together with the assumption that the linear interpolation between subsequent samples is a sufficiently accurate reconstruction of the movement of the object between two consecutive location samples. In the language of GIS, therefore, a trajectory is represented as a LINESTRING feature together with an attribute representing time.

When comparing trajectories, several aspects might get different treatment: we can use the temporal information in an absolute manner looking for similar trajectories happening at the same time (e.g., detecting when two people are following the same trajectory at the same time, enabling for example real-time ride sharing) or in a relative manner (e.g., identifying trajectories with a similar speed pattern). In contrast, we can also completely ignore the temporal information, for example, in applications such as map reconstruction in which the spatial and topological aspects of trajectories are more important.

From a formal point of view, trajectories can be represented as continuous map T from the unit interval $[0, 1]$ to space. In this setting, the unit interval represents the time along the trajectory, so given a trajectory $T : [0, 1] \to R^2$, the time 0 is mapped to the first point $t(0)$ of the trajectory and $t(1)$ points to the last point. From this perspective, we can see that the set of all continuous trajectories forms an infinite-dimensional vector space and - therefore - infinitely many sensible distance measures do exist. This is realized in a number of trajectory distances that have been defined and used throughout literature including, but not limited to, Euclidean distance, Hausdorff distance, closest points, dynamic time warping, edit distance on real trajectories (EDR), or edit distance with real penalties (ERP). See [5] for a nice introduction into various trajectory distances. Still, one of the the most classical definition from mathematics has been introduced by Fréchet [7] and can be seen as an extension to the Hausdorff distance of sets.

The Fréchet distance is best introduced informally: imagine a dog and an owner each moving along his own trajectory. They start at the beginning, have to travel through all of the trajectory both ultimately reaching the end of their trajectories and are not allowed to go backwards at any time. With these rules in place, the Fréchet distance is the minimum length of a leash connecting the dog and his owner, where the minimization has to be taken over all possible movements. More formally,

$$d_F(T_1, T_2) = \inf_{\alpha,\beta} \sup_{t \in [0,1]} ||T_1(\alpha(t)) - T_2(\beta(t))||$$

defines the Fréchet distance of T_1 and T_2, where α and β range over all possible continous and non-decreasing functions $\alpha, \beta : [0, 1] \to [0, 1]$. First note that this distance measure is very good in that it fulfills all properties of a metric, especially, the triangle inequality and that it is exploiting the temporal evolution of the trajectory without fixing it too much. In the real world, for example, two trajectories from the same road will have a very small Fréchet distance even in different traffic conditions. One of the main drawbacks of Fréchet distance is that it is difficult to compute both from a computational complexity perspective as well as from an implementation point of view in that it is not easy to get all details right and efficient on real computers. In this context, the aim of the challenge is twofold: First, we want to show that even though the worst-case complexity is quadratic, it is possible to exploit spatial indexing structures in case of realistic mobility trajectories in order two answer hundreds of queries per second on large datasets. Second, we wanted the community to provide highly performant reference implementations which can be used in GIS research making the application of Fréchet distance easier in practice.

2 Sponsors and Supporters

This year, the challenge was sponsored by NVIDIA and IBM and it is a great help for the SIGSPATIAL community. NVIDIA sponsored the challenge by donating a current NVIDIA Titan X GPU, which we hope will trigger some research on GPU computing in the spatial domain. IBM supported our activities by donating $300 in cash for the second place and $200 in cash for the third place. Additionally, the authors of the best three submissions were invited to write a short paper about the key ideas of their approaches and to give an oral presentation of their work in a conference special session.

3 Problem Description

The participants of the challenge were given a large set of map-matched trajectories in the San Francisco area. These trajectories consist of shortest paths and fastest paths according to travel time each with and without some random restrictions in the street network. A sample of this dataset is depicted in Figure 1. This dataset contains 20,199 trajectories of varying length rangin from 10 points per trajectory to 768 points with a mean length of 247 points. The dataset contains roughly five million points in total.

The background for generating a dataset of this type is the fact that the structure of quite short and efficient paths is dominant in real-world trajectory datasets and greatly simplifies the practical complexity of calculating the Fréchet distance. Similarly, we created a set of 30,000 random queries with distances drawn from a uniform random distribution bound by the dataset diameter.

Figure 1: The GIS Cup Dataset

Given this data, the authors were invited to provide a program that reads in a dataset and a set of queries and computes a multitude of Fréchet range queries at once. As we expect that no strongly subquadratic algorithm for computing the Fréchet distance exists [3], we thought that the challenge would be solved through smart spatial indexing with only a minor influence of the quality of the Fréchet distance decision algorithm. Given that the number of distance calculations for a naive solution calculating the distance matrix of all trajectories would need 605,970,000 trajectory comparisons, we expected sophisticated spatial pruning to be the key to this problem.

4 Submissions and Winners

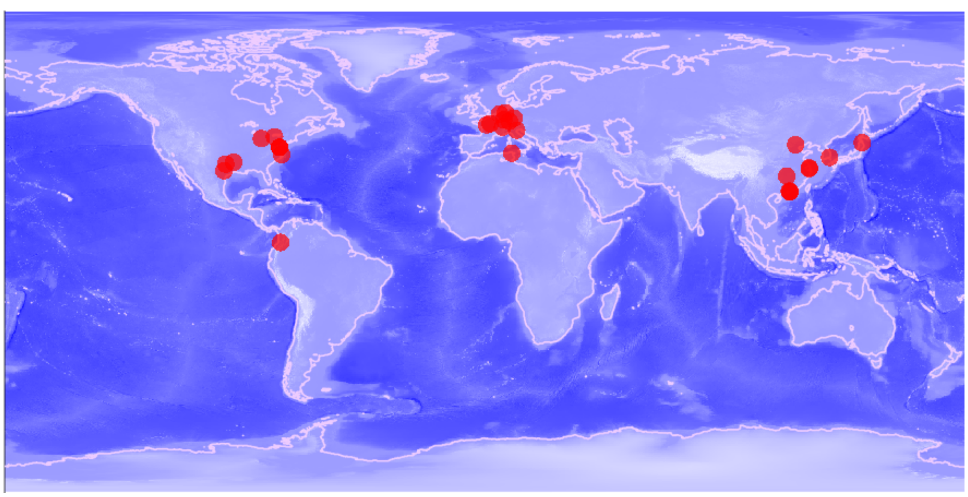

Figure 2: Overview of submissions

The challenge received considerable attention with 28 submissions from 69 individual authors and contributors from all over the world. The spatial distribution of authors as depicted in Figure 2 is quite similar to the distribution of participants in ACM SIGSPATIAL GIS and shows, therefore, that the challenge has attracted a representative subset of our community. Interestingly, there is a strong cluster in central Europe including Germany and the Netherlands. This is possibly due to the fact that there is a long tradition in studying the Fréchet distance in this area following up on seminal work by Helmut Alt, who introduced the free-space diagram for computing the Fréchet distance in a joint work with Michael Godau [1].

The best three submissions were written in C++ and used some spatial indexing based on the first and last point of the candidate trajectory exploiting that the leash will need to connect these two pairs of points in any valid reparametrization. Then, they differ on how to further reduce the set of computations by identifying true positives and true negatives with simple computations. Finally, the free-space diagram is being solved for the remaining candidates with different implementations. For more details, the reader is referred to the excellent papers invited for the three best submissions. In additon, it is great to see that all source code has been published, too – the papers contain links to the implementations.

The submission of Bringman and Baldus wins the cup for being fastest in the evaluation. They first uses some coarse spatial index, then some heuristics to prune further, and finally a recursive free space decision algorithm for validating the remaining candidates [2]. The second place was won by the submission of Buchin, Diez, van Diggelen and Meulemans [4]. They first uses a grid index to manage the trajectory endpoints and creates up to four simplifications of the trajectories. The problem is then first solved on the simplifications proceeding from coarse to fine finally solving the full Fréchet decision problem on the original pair of trajectories. The third place is given by the submission of Fabian Dütsch and Jan Vahrenhold [6]. They introduce a concept called the annulus of a trajectory and apply some other heuristics for efficiently pruning candidates. Finally, the Fŕechet decision problem is being solved. This submission was only slightly slower than the preceding two submissions and notably stable even with malformed datasets.

Acknowledgments

We want to acknowledge the support from our sponsor NVIDIA for donating an NVIDIA Titan X GPU to the winner of this challenge and our sponsor IBM for donating $300 to the second place and $200 to the third place. Furthermore, we want to express our thanks to the organizing committee of the ACM SIGSPATIAL Conference on Advances in Geographic Information Systems for hosting this competition.

References

[1] H. Alt and M. Godau. Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5:75–91, 1995.

[2] J. Baldus and K. Bringmann. A fast implementation of near neighbors queries for fréchet distance. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2017.

[3] K. Bringmann. Why walking the dog takes time: Frechet distance has no strongly subquadratic algorithms unless seth fails. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 661–670. IEEE, 2014.

[4] K. Buchin, Y. Diez, T. van Diggelen, and W. Meulemans. Efficient trajectory queries under the fréchet distance. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2017.

[5] K. Deng, K. Xie, K. Zheng, and X. Zhou. Trajectory indexing and retrieval. *Computing with spatial trajectories*, pages 35–60, 2011.

[6] F. Dütsch and J. Vahrenhold. A filter-and-refinement-algorithm for range queries based on the fréchet distance. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2017.

[7] M. M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22(1):1–72, 1906.