

(k, ℓ) -Medians Clustering of Trajectories Using Continuous Dynamic Time Warping

Milutin Brankovic*
University of Sydney
Darlington, NSW, Australia
mbra7655@uni.sydney.edu.au

Kevin Buchin*
TU Eindhoven
Eindhoven, Netherlands
k.a.buchin@tue.nl

Koen Klaren*
TU Eindhoven
Eindhoven, Netherlands
k.r.klaren@student.tue.nl

André Nusser*
Max Planck Institute for Informatics
Graduate School of Computer Science
Saarbrücken, Germany
anusser@mpi-inf.mpg.de

Aleksandr Popov*
TU Eindhoven
Eindhoven, Netherlands
a.popov@tue.nl

Sampson Wong*
University of Sydney
Darlington, NSW, Australia
swon7907@uni.sydney.edu.au

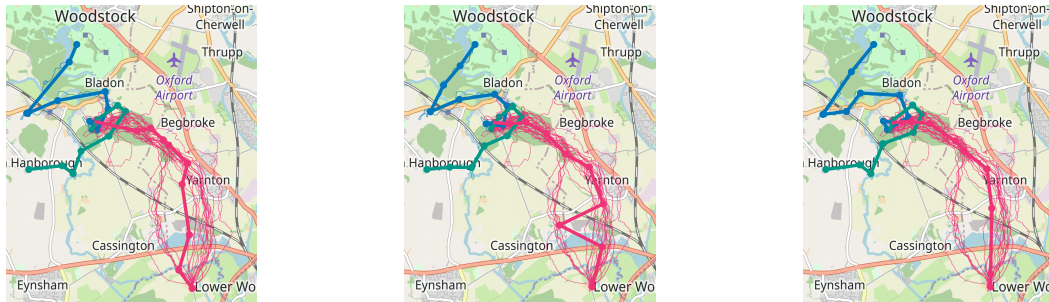


Figure 1: Clustering using the Fréchet distance (left), dynamic time warping (middle), and our new approach called continuous dynamic time warping (right), where the color denotes the clusters and the bold trajectories are the cluster centers. While the Fréchet distance shows a strong influence of outliers and dynamic time warping shows discretization issues, clustering via continuous dynamic time warping gives arguably the most natural results. Map data © OpenStreetMap contributors. [17]

ABSTRACT

Due to the massively increasing amount of available geospatial data and the need to present it in an understandable way, clustering this data is more important than ever. As clusters might contain a large number of objects, having a representative for each cluster significantly facilitates understanding a clustering. Clustering methods relying on such representatives are called *center-based*. In this work we consider the problem of center-based clustering of trajectories.

In this setting, the representative of a cluster is again a trajectory. To obtain a compact representation of the clusters and to avoid overfitting, we restrict the complexity of the representative trajectories by a parameter ℓ . This restriction, however, makes discrete distance measures like dynamic time warping (DTW) less suited.

There is recent work on center-based clustering of trajectories with a continuous distance measure, namely, the Fréchet distance. While the Fréchet distance allows for restriction of the center complexity, it can also be sensitive to outliers, whereas averaging-type distance measures, like DTW, are less so. To obtain a trajectory clustering algorithm that allows restricting center complexity and is more robust to outliers, we propose the usage of a continuous version of DTW as distance measure, which we call *continuous dynamic time warping (CDTW)*. Our contribution is twofold:

- (1) To combat the lack of practical algorithms for CDTW, we develop an approximation algorithm that computes it.
- (2) We develop the first clustering algorithm under this distance measure and show a practical way to compute a center from a set of trajectories and subsequently iteratively improve it.

To obtain insights into the results of clustering under CDTW on practical data, we conduct extensive experiments.

CCS CONCEPTS

- Information systems → Geographic information systems;
- Theory of computation → Computational geometry.

KEYWORDS

Trajectory Similarity, Trajectory Clustering, Continuous Dynamic Time Warping.

* Authors listed in alphabetical order. All authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGSPATIAL '20, November 3–6, 2020, Seattle, WA, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8019-5/20/11...\$15.00

<https://doi.org/10.1145/3397536.3422245>

ACM Reference Format:

Milutin Brankovic, Kevin Buchin, Koen Klaren, André Nusser, Aleksandr Popov, and Sampson Wong. 2020. (k, ℓ) -Medians Clustering of Trajectories Using Continuous Dynamic Time Warping. In *28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '20)*, November 3–6, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3397536.3422245>

1 INTRODUCTION

As smartphones have become a ubiquitous part of our daily lives, access to GPS technology has become more convenient than ever. In general, the prevalence of tracking technologies have led to an abundance of GPS data. For example, researchers can track animals in large quantities to help study their behavior, enabling new scientific insights in the field of behavioral biology.

With the blessing of easily available GPS data comes the burden of analyzing large quantities of data, in particular, the need for suitable methods to extract relevant information from the collected data. A classical approach to make large data sets understandable is to cluster the data into different groups. While this makes it easy to identify different groups in the data, it may still be simply too much information to digest or to visualize. Identifying a good representative for each group helps condense the information even further. Clustering methods that rely on such representatives—also called cluster centers—are referred to as *center-based clustering*.

Let a trajectory be represented by a polygonal chain in Euclidean space. While point sets in Euclidean space have a canonical distance measure (the L_2 -norm), for trajectories there is no such clear choice. Thus, trajectory clustering has been considered under several different distance measures. Two popular distance measures are Dynamic Time Warping (DTW) and the Fréchet distance, but each have their own shortcomings.

Dynamic Time Warping (DTW) matches the vertices from one trajectory to vertices on the other such that the summed distance between matched vertices is minimized. In DTW, only vertices of the trajectory are considered, and the edges between these vertices are ignored. As such, DTW is sensitive to the relative sampling rates of the trajectories and provides a poor matching between a high-complexity trajectory and a low-complexity trajectory. See Figure 2 for an example of this issue. For this reason, DTW is inappropriate for computing low-complexity cluster centers for high-complexity trajectories, as shown in Figure 3.

The Fréchet distance captures the minimal cost of a continuous deformation of one trajectory into another, where the cost of deformation is the maximum distance between a point and its transformed counterpart. Recent work on center-based clustering of trajectories has used the Fréchet distance [2, 3, 6]. The differences of DTW and the Fréchet distance are, first, that the Fréchet distance considers both edges and vertices of the trajectory, and second, that the Fréchet distance is a bottleneck measure (i.e., its cost is a maximum) whereas DTW is a summed measure. As a result of the Fréchet distance being a bottleneck measure, a shortcoming of using the Fréchet distance to compute cluster centers is that it may be sensitive to outliers in the trajectory data. See Figure 4 for an example where this issue occurs.

To overcome the shortcomings of DTW and the Fréchet distance for center-based trajectory clustering, we propose the usage of a

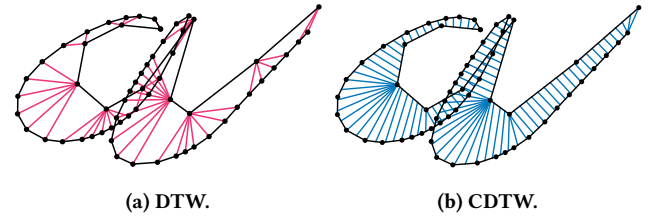


Figure 2: An example of discrete and continuous alignments of points along trajectories of differing complexities by DTW and CDTW, respectively.

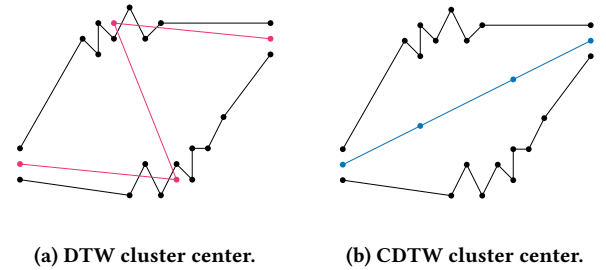


Figure 3: An example where the low-complexity DTW cluster center is inappropriate for high-complexity trajectories.

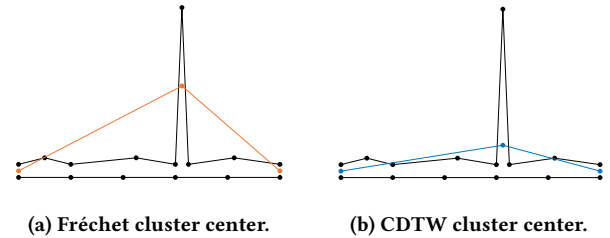


Figure 4: An example where the quality of the Fréchet cluster center deteriorates in the presence of an outlier.

continuous version of DTW. We call this measure the *continuous dynamic time warping distance (CDTW)* and formally define it in Section 3. The advantages of computing cluster centers with CDTW are that it is appropriate for low-complexity centers (see Figure 3) and is robust to outliers (see Figure 4). In Section 6, we provide empirical evidence in support of these claims.

The distance measure we call CDTW was originally introduced as the summed or average Fréchet distance by Buchin [5, Ch. 6]. A different distance measure with the name CDTW was introduced by Munich and Perona [16]. This measure is continuous in the sense that a vertex of a trajectory may be aligned to any point on the other trajectory, instead of only to another vertex. It is, however, still based on discrete summation over the vertices in the same manner as DTW and is therefore still prone to issues regarding sampling of the trajectories. A general definition for a continuous distance measure for trajectories using path integrals with the name CDTW

was given by Efrat et al. [8]. They also sketch an approximation algorithm for a special case of our definition of CDTW, but it is based on numerical methods for solving eikonal equations, rendering it impractical. A $(1 + \varepsilon)$ -approximation algorithm for our definition of CDTW was introduced by Maheshwari et al. [14]. Its running time is upper-bounded by $O(\zeta^4 n^4 / \varepsilon^2)$, where n is the complexity of the trajectories and ζ is the maximal ratio between the lengths of any pair of segments from the trajectories. This dependency on the ratio between segments makes the algorithm unsuited for the problem of (k, ℓ)-center based clustering, where one trajectory consists of many short segments and the other of a few long segments. As far as we are aware, there are no exact algorithms known for computing the CDTW distance.

For point sets, clustering algorithms that are parametrized by the number of clusters, like k -center or k -means clustering, are very common. Trajectories, however, pose the problem that their complexity can vary. Increasing the complexity of the centers also increases the adaptability of the centers to the data at the cost of overfitting and complicated centers. Therefore, recent work introduced the notion of (k, ℓ)-clustering for trajectories [2] (and there has been a previous closely related notion for time series [6]), which is a center-based clustering where we require the clustering to have k clusters and each center can have complexity at most ℓ . More precisely, we consider the following two variants of (k, ℓ)-clustering:

- (1) (k, ℓ)-center clustering: Find k centers of complexity at most ℓ that minimize the maximal distance that any trajectory has to its closest center.
- (2) (k, ℓ)-medians clustering: Find k centers of complexity at most ℓ that minimize the sum of distances of each trajectory to its closest center.

See Section 5 for a formal definition and detailed discussion. Interestingly, different notions of clustering seem more natural for different distance measures. As the Fréchet distance and (k, ℓ)-center clustering minimize the maximal distance, it seems natural to combine them. Similarly, as DTW / CDTW and (k, ℓ)-medians minimize a sum / integral, it again seems natural to combine them.

An approximation algorithm for (k, ℓ)-center clustering using the continuous Fréchet distance was given by Buchin et al. [2]. This algorithm, with additional practical modifications, was later implemented and applied to real-world data by Buchin et al. [4]. For the discrete Fréchet distance, Buchin et al. [3] designed approximation algorithms for (k, ℓ)-medians clustering and (k, ℓ)-center clustering. Finally, Petitjean et al. [18] introduced an algorithm for computing low-complexity centers for a cluster of trajectories using DTW distance.

2 CONTRIBUTIONS

Our main contribution is a trajectory clustering implementation using the CDTW distance measure. The advantages of this measure over existing measures are that it is appropriate for low-complexity centers and robust to outliers. To the best of our knowledge, our implementation is the first to cluster trajectories using CDTW. The two main contributions to enable such an implementation are:

- As far as we are aware, there are no exact algorithms for CDTW, and the existing approximation algorithms have a high dependency on both the complexity and the lengths

of the input trajectories [14]. We overcome this obstacle by providing a practically efficient additive approximation algorithm for computing the CDTW distance, and an efficient implementation of said algorithm. See Section 4 for details.

- Again, as far as we are aware, there are no known algorithms for (k, ℓ)-medians clustering of trajectories under CDTW. We overcome this obstacle by following the method of Buchin et al. [2]. This involves computing an initial (k, ℓ)-clustering by combining the Gonzalez algorithm or PAM with a trajectory simplification algorithm, and then improving the cluster centers using methods similar to those found in the practical work by Buchin et al. [4]. We explore several methods for improving cluster centers, and perform extensive experiments to compare these methods. See Section 5 for details.

Our experimental results show that our method, which clusters trajectories under the CDTW measure, outperforms similar methods which cluster trajectories under either the Fréchet distance or the DTW measure. The improvement of the CDTW measure over the DTW measure is particularly noticeable for low-complexity cluster centers, and over the Fréchet distance in the presence of outliers. See Section 6 for details.

3 CONTINUOUS DYNAMIC TIME WARPING

We represent trajectories as polygonal chains, i.e., series of connected line segments. A trajectory P is given by the sequence of its vertices (p_1, \dots, p_n) , where each $p_i \in \mathbb{R}^d$ for some $d \geq 1$. The DTW distance and the CDTW distance are summations over distances between aligned points along the trajectories. Among the different possibilities for distance measures between points, we consider the (squared) Euclidean distance to be the most natural choice. Both regular and squared Euclidean distances are commonly used in DTW computation. Furthermore, methods for center computation, such as the least-squares method for linear regression as well as k -means clustering, also use the squared Euclidean distance. Thus, in the remainder we also focus on the squared Euclidean distance measure, i.e., the distance between two points $p, q \in \mathbb{R}^d$ is defined as $d(p, q) = \|p - q\|_2^2$.

A central definition for the DTW distance is the notion of a *discrete warping path*. Given a pair of trajectories $P = (p_1, \dots, p_n)$ and $Q = (q_1, \dots, q_m)$, define $W = (w_1, \dots, w_k)$ with $w_i \in \{1, \dots, n\} \times \{1, \dots, m\}$. This warping path aligns vertices in P with vertices in Q , meaning that $w_k = (i, j)$ corresponds to a match of vertex $p_i \in P$ to vertex $q_j \in Q$. The following restrictions are commonly imposed on discrete warping paths:

- (1) The first and last vertex in P must be matched to the first and last vertex in Q , respectively.
- (2) Each vertex in P must be matched to at least one vertex in Q , and vice versa.
- (3) The alignment may not move backwards on the trajectories. More precisely, the indices in the warping path must be monotonically increasing, i.e., $i_{k-1} \leq i_k$ and $j_{k-1} \leq j_k$.

Note that from (2) and (3) it follows that the warping path can only step to neighboring vertices. For DTW, the cost of a warping path is the sum of distances between aligned vertices. The DTW distance $\delta_D(P, Q)$ between a pair of trajectories is then the value

of the minimal-cost warping path:

$$\delta_D(P, Q) := \min_W \sum_{(i,j) \in W} d(p_i, q_j).$$

Let $L(P)$ denote the total arc length of a trajectory P . We define the parameter space $\mathbb{P} = [0, L(P)] \times [0, L(Q)]$ as an axis-aligned rectangle representing all pairs of points in trajectories P and Q . Let $P(t)$ denote the point that lies arc length t along P . Then, each $(p, q) \in \mathbb{P}$ corresponds to a pair of points $(P(p), Q(q))$. We define a height function $h: \mathbb{P} \rightarrow \mathbb{R}^+$ over the parameter space that gives the distance between the corresponding points, i.e., $h((p, q)) := d(P(p), Q(q))$. Figure 5 shows an example of the parameter space and height function for the pair of trajectories in Figure 2.

A *continuous warping path* $\pi: [0, 1] \rightarrow \mathbb{P}$ aligns points in P to points in Q , i.e., if (p, q) is in the image of π , the point $P(p)$ is aligned to the point $Q(q)$. We impose restrictions on continuous warping paths similar to those for its discrete counterpart.

- (1) The first and last vertex in P must be matched to the first and last vertex in Q , respectively.
- (2) Every point in P must be matched with at least one point in Q , and vice versa.
- (3) The alignment may not move backwards over the trajectories. More precisely, both coordinates of π must be monotonically increasing.

It follows that π is a continuous function that monotonically increases from $\pi(0) = (0, 0)$ to $\pi(1) = (L(P), L(Q))$. For CDTW, the cost of a warping path is the integral over distances between aligned points. Therefore, a natural definition is the path integral over π through h :

$$\int_{\pi} h(r) dr = \int_0^1 h(\pi(t)) \cdot \|\pi'(t)\| dt.$$

It may be desirable to use different norms for the speed of the path $\|\pi'(t)\|$, so we generalize the definition to support this. The CDTW distance between a pair of trajectories using the L_p -norm in parameter space is the cost of the continuous warping path with minimal cost:

$$\delta_{CD}^p(P, Q) := \min_{\pi} \int_0^1 h(\pi(t)) \cdot \|\pi'(t)\|_p dt.$$

By using $p = 1$, the path length $\int_0^1 \|\pi'(t)\|_1 dt = L(P) + L(Q)$ is the same for all warping paths, as noted by Buchin [5], thus allowing for normalization and easier comparison. In a similar setting, $p = \infty$ was used to place a speed limit on the warping path [20]. We identify $p = 1$ and $p = \infty$ as good choices for p , in particular due to their natural meaning as the sum and the maximum of speeds along the trajectories. In this work we use $p = 1$ and thus define $\delta_{CD}(P, Q) := \delta_{CD}^1(P, Q)$. An example of minimal-cost warping paths for both DTW and CDTW is shown in Figure 5.

4 COMPUTING CDTW

The task of computing the continuous dynamic time warping distance reduces to finding an optimal warping path, i.e., a warping path for which the CDTW cost is minimized. We first split the problem into smaller subproblems which can easily be solved exactly, and then apply existing algorithms to combine the exact solutions into an approximation of an optimal warping path.

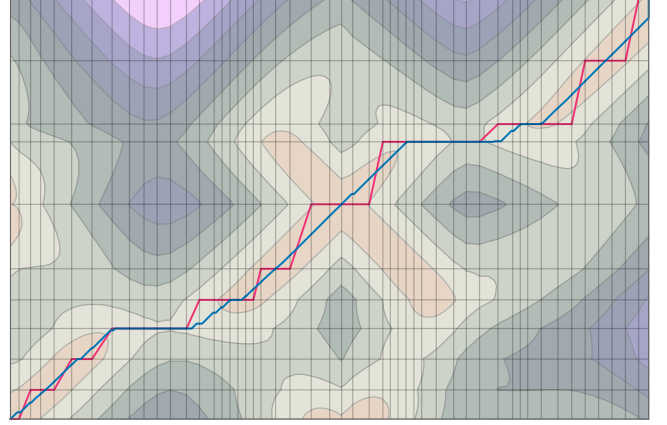


Figure 5: Warping paths of DTW and CDTW through parameter space corresponding to the trajectories in Figure 2. The grid lines correspond to the vertices of the trajectories.

As follows from the description in Section 3, the segments of the trajectories induce a grid of axis-aligned rectangular cells over the entire parameter space, such that each cell uniquely corresponds to a pair of segments. See Figure 5 for an example of such a parameter space with its grid of cells. The height function over a single cell is therefore a function over linear interpolations of the segments. More precisely, we define the height function at the point $(p, q) \in \mathbb{P}$ in a parameter space cell as

$$h((p, q)) := (p - a)^2 + (q - b)^2 + 2\lambda(p - a)(q - b) + c,$$

where the values for $a, b \in \mathbb{R}$ are based on the offset of the cell's corresponding segments along the infinite lines on which they lie. Note that these values are fixed for each cell. The value of $-1 \leq \lambda \leq 1$ is the dot product between the unit direction vectors of the segments and is fixed for each cell as well. The value of c is the smallest distance between the infinite lines of the segments and is therefore non-zero only in the degenerate case of parallel lines. The *level sets* of the height function h form concentric ellipses with center (a, b) and eccentricity based on λ , possibly degenerating into parallel strips. Let ℓ_m be the line through the point (a, b) with slope 1, coinciding with the xy -monotone axes of the ellipses. Figure 6 shows a number of these level sets and the ℓ_m line.

Given a cell C and points $s, t \in C$ with $s_x \leq t_x$ and $s_y \leq t_y$ and assuming without loss of generality that s and t are the bottom left and top right corners of C , respectively, we can compute the optimal warping path from s to t as shown by Maheshwari et al. [14]:

- If ℓ_m intersects the border of C , let c_s and c_t be the points of intersection closest to s and t , respectively. Then the optimal warping path is (s, c_s, c_t, t) .
- Otherwise, let c be the corner point closest to ℓ_m . Then the optimal warping path is (s, c, t) .

An example for both cases is shown in Figure 6. Note that in both cases the optimal warping path consists of a constant number of linear pieces. We integrate along each linear piece to compute the cost of the warping path from s to t .

With this algorithm for computing the optimal warping path through single cells, it remains to find the points where the optimal

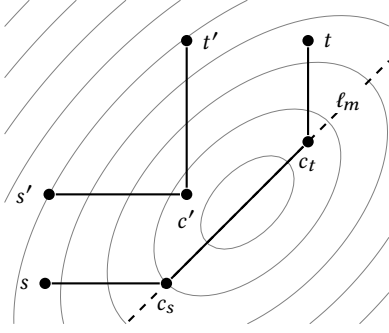


Figure 6: Two types of optimal warping paths through a single cell: (s, c_s, c_t, t) via ℓ_m , and (s', c', t') missing ℓ_m .

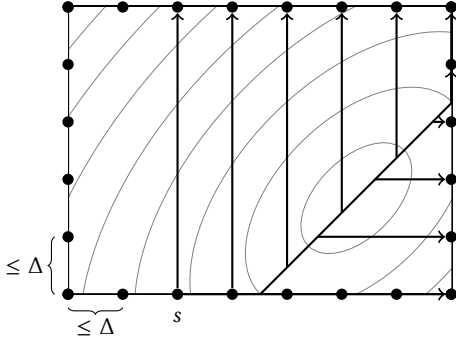


Figure 7: Example placement of vertices along the borders of a cell, and an example of the edges originating from one such vertex s visualized by the optimal warping paths.

warping path crosses from one cell to another. To this end, define a graph over the parameter space. Take the corner points of the cells as the vertices of the graph. To improve the accuracy of the warping path, we uniformly sample additional vertices (called Steiner points) along the borders of each cell, such that the distance between two neighboring sampled vertices is at most Δ in the original Euclidean space. We add an edge between two vertices s and t if both vertices lie in the same cell and are positioned such that $s_x \leq t_x$ and $s_y \leq t_y$. Each edge is assigned a weight equal to the cost of the optimal warping path between its incident vertices. An example of vertex placement and edges is shown in Figure 7. The cost of the shortest path through this graph from $(0, 0)$ to $(L(P), L(Q))$ is an additive approximation to the cost of the optimal warping path.

Any algorithm that finds the shortest path between two given vertices in an acyclic graph with positive edge weights will be sufficient in finding the approximation of the optimal warping path. In our implementation we use bidirectional Dijkstra’s algorithm as it often needs to consider fewer vertices than regular Dijkstra’s algorithm. This algorithm searches from both the start and goal vertex until the two search fronts meet with an appropriate stopping condition. In practice we do not compute the entire graph in advance; instead, for a given vertex, we compute the adjacent vertices and incident edges on the fly as they are explored by the shortest path algorithm.

5 CLUSTERING ALGORITHMS

Before we present our clustering algorithms, we first formally define the two notions of clustering that were presented in Section 1. Let C be a set of trajectories and let k and ℓ be positive integers. Let d be some distance measure for trajectories. In a (k, ℓ) -clustering problem, we seek to compute a set of k center trajectories Q of complexity at most ℓ which is optimal according to some cost function. The two notions we consider are:

- (1) In (k, ℓ) -center clustering, the maximum distance from each trajectory in C to the nearest center in Q is minimized, i.e., the relevant cost function is

$$\phi_\infty(Q) := \max_{C \in C} d(C, Q).$$

- (2) In (k, ℓ) -medians clustering, the sum of distances from each trajectory in C to the nearest center in Q is minimized, i.e., the relevant cost function is

$$\phi_1(Q) := \sum_{C \in C} d(C, Q).$$

For some variants we also consider (k, ℓ) -medians clustering where we use the squared L_2 -norm as the underlying distance measure. This can also be considered a means clustering.

Our high-level strategy for (k, ℓ) -medians clustering is as follows. We first compute an initial set of k centers, each of them being an ℓ -simplification of a trajectory from the input set. We then apply an iterative center improvement algorithm to obtain cluster centers that are not restricted to the vertices of the input trajectories.

5.1 Computing the Initial Clustering

We use the *partitioning around medoids (PAM)* method, also called k -medoids, proposed by Kaufman and Rousseeuw [12, 13, Ch. 2] and recently improved by Schubert and Rousseeuw [23]. However, we adapt the method to the (k, ℓ) -medians clustering setting. The PAM method for k -medians clustering in general metric spaces works by first greedily choosing an initial set of centers. In this step, we compute the reduction in the sum of distances from trajectories to cluster centers for each trajectory and then choose the one that gives the biggest reduction as a new center; we add k centers this way. Next, the algorithm enters a local search phase where for each center c_i , we evaluate the possible reduction in the sum of distances if we swap c_i for a non-center x . We perform the swap that gives the biggest difference and repeat the procedure iteratively until we find a local optimum. Adapting this approach to our setting is straightforward. Whenever we consider a trajectory as a center, we compute the distances to its ℓ -simplification.

We also adapt the *greedy algorithm of Gonzalez* to the setting of (k, ℓ) -clustering [9]. The idea is simple: we start with a random input trajectory as the first center; then we iteratively pick the trajectory which is furthest from the already picked centers until we get k centers. The only difference is that we pick the ℓ -simplifications of the center trajectories, and compute the distances accordingly. Even though the Gonzalez algorithm optimizes for the k -center cost function, it is known to be a $2n$ -approximation for the k -medians problem, where n is the number of input points [10, Ch. 4]. Therefore, we use the Gonzalez algorithm as a baseline for our initial clustering approaches. We also consider an alternative

clustering algorithm that first computes a (k, ℓ) -clustering using the Gonzalez algorithm and then applies the local search of PAM.

5.2 Trajectory Simplification

To use the approach described in Section 5.1, we need to compute the ℓ -simplifications of the center trajectories when computing an initial clustering. The choice of simplification approach to use is somewhat independent of the clustering methods: we need to choose an approach that allows us to restrict the maximum complexity of the simplified trajectory to ℓ and gives good simplifications, both visually and in terms of CDTW. As this is only used as the initial step before improving the centers, we choose the simpler approach of vertex-restricted simplification, where the vertices of the simplified trajectory are a subset of the vertices of the original trajectory. We focus on two simplification methods: the dynamic programming approach suggested by Imai and Iri [11] and the greedy approach by Agarwal et al. [1].

The greedy approach is very efficient, offers reasonable guarantees for the Fréchet distance, and works well in practice. The main idea is: given a threshold and a starting vertex of the trajectory p_i , go along the trajectory until we reach a vertex p_j such that the distance of choice between the segment $p_i p_{j+1}$ and the subtrajectory from p_i to p_{j+1} exceeds the threshold. Here i and j are indices of vertices of the trajectory with $j > i$. Once that happens, we add the segment $p_i p_j$ to our simplification and set the vertex p_j as the next starting point. We repeat the process until we reach the end of the trajectory. This method can be used to obtain a simplification of complexity ℓ by performing a binary search on the threshold value.

The Imai–Iri approach allows to compute an optimal simplification in a local sense, i.e., assuming the matching is fixed at the points that coincide between the trajectory and its simplification. There are two reasonable approaches to its implementation:

- (1) Given a threshold, save all the possible line segments between vertices of the trajectory for which the distance to the corresponding subtrajectory is below the threshold. Then find the shortest path from the beginning to the end of the trajectory. This can be adapted to targeting the specific ℓ using binary search on the threshold value.
- (2) Given a target ℓ , use a dynamic program to record the costs of all possible shortcuts in ℓ steps and find the lowest-cost solution. This approach seems more suitable for sum-based distances like DTW and CDTW.

It is not obvious which approach would be preferable in our setting, since any of these could be used with DTW, the Fréchet distance, and CDTW. We conduct an experimental evaluation and describe the results in Section 6.3. Based on these results, we choose the greedy approach with Fréchet distance in all our clustering pipelines.

5.3 Improving Cluster Centers

Petitjean et al. [18] introduced a method called *DTW Baricenter Averaging (DBA)* to improve the centers in k -means clustering using the DTW distance. Let C be a trajectory cluster and let $C = (c_1, \dots, c_\ell)$ be its initial center. The DBA method first computes the DTW matching between the center C and all trajectories in the cluster C . For all $1 \leq i \leq \ell$, let \tilde{c}_i be the average of all points matched to c_i . We replace the initial center C with the new center

$\tilde{C} = (\tilde{c}_1, \dots, \tilde{c}_\ell)$ if it induces a lower cost. The procedure is repeated until the new center does not improve upon the current center.

We use a similar method to improve the cluster centers arising from the CDTW distance. We call our method *CDTW Barycenter Averaging (CDBA)*. A key difference between the DBA method and our CDBA method is that ours uses a continuous matching, so a vertex on C may be mapped to an entire subtrajectory rather than a set of vertices. In this situation, we uniformly sample points on the matched subtrajectory. The number of sampled points is proportional to the ratio of the length of the subtrajectory to the length of the trajectory. See Figure 8 for an example where the matching is sampled to compute the updated centers.

The sampling method for vertices mapped to a continuous subtrajectory has been previously used by Buchin et al. [4]. Their work focuses on (k, ℓ) -center clustering using the Fréchet distance, and we refer to their method as *free space averaging (FSA)*. In the FSA method, rather than updating vertices of the center trajectory to be the mean of the points they are matched to, vertices are assigned to the center of the minimum enclosing circle of the points they are matched to in the Fréchet matching.

The intuition behind the center update methods described above is that by moving vertices of the center trajectory closer to the points on cluster trajectories to which they are aligned with respect to the distance measure under consideration, we expect the new center trajectory to be a better fit for the cluster. This intuition makes the implicit assumption that the alignment between the center trajectory and trajectories in the cluster does not change too much after the update. The update method for each distance function also depends on the type of clustering considered to be most natural for that particular distance function. The Fréchet distance minimizes the maximum distance between aligned points. Similarly, k -center clustering minimizes the maximum distance between each trajectory and the trajectories in the associated cluster, and so k -center clustering seems a natural choice for clustering with the Fréchet distance. The FSA method described was designed by Buchin et al. [4] for the purpose of improving centers in (k, ℓ) -center clustering. Both DTW and CDTW are distance measures that minimize the sum or integral of distances between aligned points. Therefore, k -means or k -medians clustering seem to be more natural choices for these distance measures. The DBA center update method we have described was designed by Petitjean et al. [18] in the context of k -means clustering under the DTW distance measure, while we apply the CDBA method for k -medians clustering under CDTW.

The CDBA method is based on the strategy of moving vertices. However, the definition of CDTW also takes into consideration the internal points on edges. Therefore, we also consider a center update method which aims to optimize the position of edges, rather than just vertices. With this in mind, we introduce an alternative center update method for (k, ℓ) -medians clustering under CDBA which we call the *wedge method*. Given an initial center trajectory $C = (c_1, c_2, \dots, c_\ell)$ and the associated cluster C , the wedge method works as follows. For each c_i , $1 < i < \ell$, we define the line segments $s_L = c_{i-1}c_i$ and $s_R = c_i c_{i+1}$. We then define the sets L and R to be all vertices of trajectories in C aligned by CDTW to s_L and s_R , respectively. For each vertex p of a trajectory $P \in C$ aligned to either s_L or s_R under CDTW, we define $w(p)$ to be the sum of lengths of the segments preceding and following p on P . We search

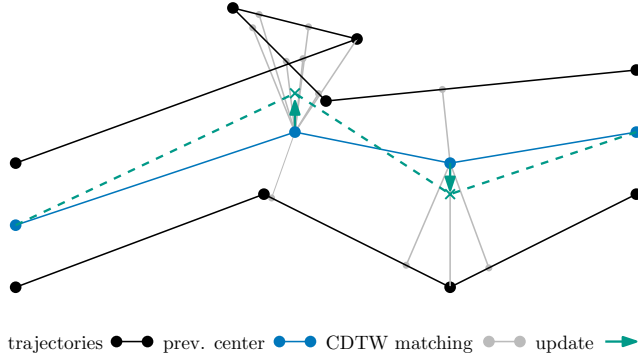


Figure 8: The CDBA update method, with the CDTW matching sampled uniformly to compute the updated centers.

for a point \tilde{c}_i by perturbing c_i so as to minimize

$$\sum_{p \in L} w(p) \cdot \gamma(p, c_{i-1} \tilde{c}_i)^2 + \sum_{p \in R} w(p) \cdot \gamma(p, \tilde{c}_i c_{i+1})^2,$$

where $\gamma(p, s)$ denotes the distance from a point p to the closest point on a line segment s .

Finally, we can perform a similar optimization for the first and last points on the initial center trajectory. Depending on the nature of the data we are working on, it may make more sense to keep the start and end points fixed, as is the case in the data set of pigeon flight paths we study. We include the weight terms $w(\cdot)$ as we expect vertices between longer segments to contribute more to the value of the CDTW distance.

6 EXPERIMENTS

In this section, we describe the experiments performed on the methods described above and their results.

6.1 Data Sets

We have conducted experiments on three different open-source real-world data sets. The first of these is a set of handwritten characters of the English alphabet [24] from the UCI Machine Learning Repository [7]. The data set contains a few hundred examples in each of twenty classes, with each class corresponding to a character that can written in a single stroke.

We also make use of the pigeon flight path trajectory data set by Mann et al. [15]. The data set consists of flight paths of pigeons released from four distinct release sites as they fly to a common return site. For each release site, the data set contains trajectories corresponding to seven or eight distinct pigeons, each of which is released approximately twenty times.

Finally, we also make use of the stork migration data by Rotics et al. [22], made available at the Movebank data repository [21]. The data set contains the movements of 35 adult storks over four years as they migrate from Africa to Europe.

As our methods become impractical for very large trajectories (some trajectories in the stork migration data set contain over 15000 points), we pre-process the data by regularly sampling points to ensure the trajectories we use have at most a few hundred vertices. The coordinates in the pigeon and stork data sets are given

as latitude and longitude. Since our methods assume trajectories in the plane, we apply projections to the data sets. We use the transverse Mercator projection for the pigeon data set and the two-point equidistant projection for the stork data set.

We conduct experiments to evaluate the three steps in our trajectory clustering implementation, that is, our initial clustering, our trajectory simplification, and our center improvement methods. We compare our results to existing approaches. In particular, we compare the quality of trajectory simplifications based on DTW and the Fréchet distance with our algorithm based on CDTW, and we compare the quality of center improvement methods based on DBA and FSA with our CDBA and wedge methods. We make the source code of our C++ implementation publicly available.¹

6.2 Initial Clustering

We apply the PAM clustering algorithm to the pigeon flight data. For each of the four release sites, we compute (k, ℓ)-clusterings, for $k \in \{1, \dots, 7\}$ and $\ell = 10$, using the following algorithms with CDTW and greedy simplification with the Fréchet distance:

- (1) The Gonzalez algorithm;
- (2) PAM after initializing the centers using the Gonzalez algorithm;
- (3) PAM after greedily initializing the centers according to the ϕ_1 cost function.

Since the Gonzalez algorithm is randomized, when clustering via algorithms (1) and (2), we take the best of five iterations as the final result. To compare the approaches, we evaluate the values of the (k, ℓ)-medians scores for the resulting clusterings. Both PAM and PAM with Gonzalez initialization consistently outperform the vanilla Gonzalez approach. This is to be expected, since the Gonzalez algorithm does not aim to optimize the (k, ℓ)-medians score. The results also indicate that PAM and PAM with Gonzalez initialization often arrive at the same final clustering.

6.3 Trajectory Simplification

As described in Section 5.2, there are many possible approaches to trajectory simplification in our setting. We focus on the greedy approach and the dynamic programming approach, using DTW, the Fréchet distance, and CDTW. We experimentally evaluate these approaches to gauge their performance both visually and with respect to CDTW. In order to do that, we sample trajectories from two of the data sets we use, the character and the pigeon data set. We additionally subsample each trajectory at regular intervals to get trajectories of complexity 50. We then use the different methods to simplify the trajectories to complexity 12 and evaluate the resulting simplifications with respect to CDTW to the original trajectories. The results are shown in Tables 1 and 2.

As can be seen from the tables, using CDTW with Imai-Iri gives the best results, as expected. Unfortunately, this is also the slowest method, several orders of magnitude slower in our experiments than approaches that use the Fréchet distance or DTW, which makes it impractical. Using the greedy approach with CDTW gives simplifications of only slightly worse quality (see Figure 9b), but still shows high computation times.

¹See <https://github.com/Mesoptier/trajectory-clustering>.

Table 1: Simplification methods on the characters data set. Reported values are computed over the data set w.r.t. CDTW between the trajectories and their simplifications.

		mean	min	max
DTW	Greedy	255.5	1.90	2001.0
	Imai-Iri	175.6	1.58	1827.4
Fréchet	Greedy	71.4	0.88	695.2
	Imai-Iri	67.4	0.69	689.5
CDTW	Greedy	68.2	0.93	665.5
	Imai-Iri	36.9	0.47	469.1

Table 2: Simplification methods on the pigeons data set (the Bladon Heath site). Reported values are computed over the data set w.r.t. CDTW between the trajectories and their simplifications.

		mean	min	max
DTW	Greedy	0.829	0.003	9.87
	Imai-Iri	0.586	0.005	4.68
Fréchet	Greedy	0.879	0.024	10.19
	Imai-Iri	0.676	0.023	6.60
CDTW	Greedy	0.883	0.008	9.14
	Imai-Iri	0.335	0.005	3.87

Other contenders show comparable (amongst each other) results on the two data sets, as can be seen from the results in both tables. They are also all much faster; so, it makes sense to examine the specific defects that arise from each approach to better understand which methods are more suitable in our setting. On the characters data set, DTW fares poorly both using Imai-Iri and the greedy approach. Figure 9a showcases a bad example for dynamic time warping with the Imai-Iri approach, compared to the performance of Imai-Iri with CDTW on the same trajectory. Due to the discrete nature of DTW, points have to be placed somewhat regularly, including on straight-line parts of the trajectory. Since ℓ is fixed, DTW-based approaches visually perform poorly on the curved parts. This is reflected in the values of CDTW. Even though DTW performs better on (part of) the pigeon data set, using it in general settings is not ideal, since there are many real-world trajectories with both long straight segments and smooth curvature that will yield bad results, e.g., plane or road vehicle trajectories.

Approaches based on the Fréchet distance fare much better on the character data set and comparably on the pigeons data set. Visual representations of the worst examples compared to Imai-Iri CDTW are shown in Figure 10. Note that here the issue is simply the lack of attention for long segments that do not quite coincide with the trajectory, something that is cost-free for the Fréchet distance, but penalized quite heavily for CDTW. Fréchet-induced simplifications, however, do concentrate the points in the curved sections of the input instead of the straight-line segments, and tend to use less points than the available number ℓ . Note that we use the simplifications of the center trajectories only as a starting point in our clustering approach before doing further center improvements.

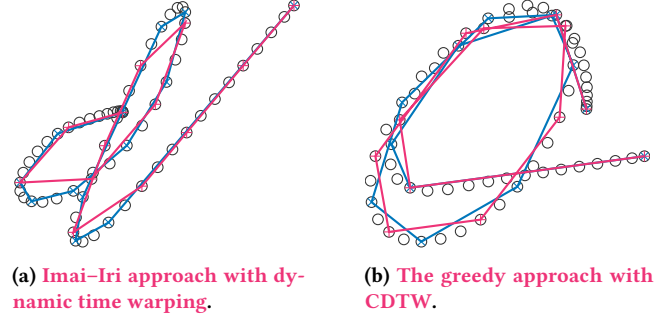


Figure 9: Example simplifications for DTW and CDTW. Over the original trajectory, two simplifications are shown in color. Imai-Iri approach with CDTW performs best.

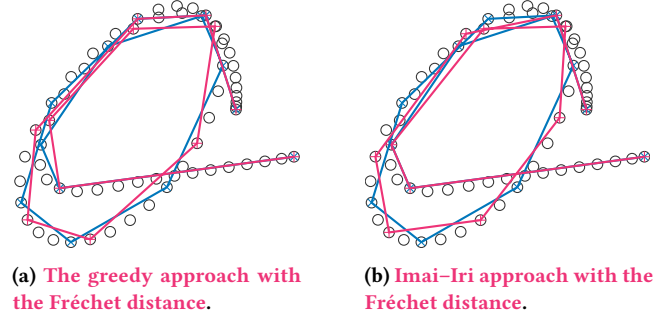


Figure 10: Example simplifications for the Fréchet distance. Over the original trajectory, two simplifications are shown in color. Imai-Iri approach with CDTW performs best.

Our center improvement strategies can add points in the places that benefit most from it, thus alleviating some of the issues with the Fréchet simplifications. So, Fréchet distance-based simplifications seem to be a good fit for our clustering approach, assuming we cannot use CDTW-based simplification due to the high running time. The greedy approach using the Fréchet distance is very fast, on par with DTW computations, so using it on the fly is feasible. Therefore, in our clustering approach we use the greedy approach with the Fréchet distance to compute the ℓ -simplifications.

6.4 Improving Cluster Centers

We conduct experiments to evaluate the quality of the DBA, FSA, CDBA, and wedge methods for improving cluster centers. First, we conduct an experiment using the character data set to compare the quality of the centers produced according to the CDTW (k, ℓ) -medians cost. In particular, since our CDBA and wedge methods are designed to minimize the CDTW (k, ℓ) -medians cost, we expect them to outperform DBA and FSA. The experiment is conducted as follows. For each character, we sample fifty trajectories from the data. For each of these sets of fifty trajectories, we compute an initial (k, ℓ) -medians clustering, for a fixed k and a range of values for ℓ , using our version of the PAM method with CDTW and greedy simplification with the Fréchet distance. We then apply four different center improvement algorithms to the same initial

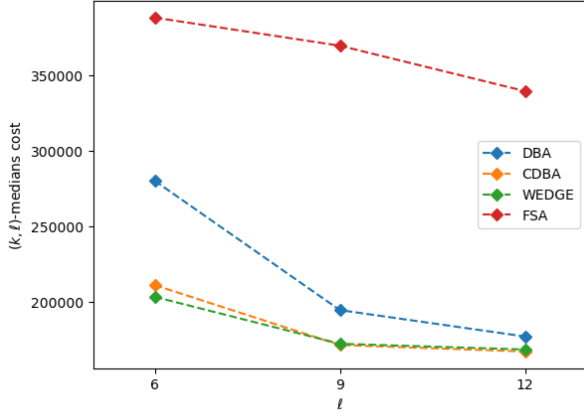


Figure 11: A comparison of the average (k, ℓ) -medians score over the twenty characters, for different values of ℓ .

clustering. Each method iteratively updates the center trajectory until the method does not make an improvement or until we reach twenty iterations.

We manually select the values of k and ℓ that reasonably reflect the underlying trajectories in the data set. For the characters data set, there seems no inherently correct value for k , so we choose $k = 2$ to accommodate for possible deviations in handwriting. Furthermore, we choose $\ell \in \{6, 9, 12\}$, as the shape of the characters can be represented by trajectories of this complexity. Note that the general problem of parameter selection is a difficult one. Reddy and Vinzamuri [19] discuss common ways to choose k . The general model selection techniques suitable for picking k (e.g., information criteria) can also be applied to choose ℓ . These methods are especially preferred for data sets where the true complexities of the underlying trajectories are difficult to estimate.

The results of the experiments for $k = 2$ and $\ell \in \{6, 9, 12\}$ are shown in Figure 11. On the y -axis are the average CDTW (k, ℓ) -medians scores over the twenty characters in the data sets. The FSA method obtains significantly higher (k, ℓ) -medians scores for all values of ℓ . The DBA method obtains lower (k, ℓ) -medians scores than FSA, but the score degrades more quickly when the complexity of the cluster center decreases. This supports our claim in the introduction that clustering based on DTW is sensitive to low-complexity cluster centers.

Next, we plot the CDTW (k, ℓ) -medians cost of the twenty character classes in Figure 12. We do so for $k = 2$ and $\ell = 6$. The costs for FSA are significantly higher than the other three methods so we omit the FSA results from the plots. As expected, the CDBA and wedge methods outperform DBA on the CDTW (k, ℓ) -medians cost on all characters. We select a few examples to highlight why we believe this is the case. In particular, we show that there are apparent visual artifacts in clusterings where DBA and FSA obtain a significantly higher (k, ℓ) -medians cost.

In Figure 13 in the appendix, FSA obtains a significantly higher (k, ℓ) -medians cost than DBA, CDBA and the wedge method. The green center computed by FSA is significantly distorted while the

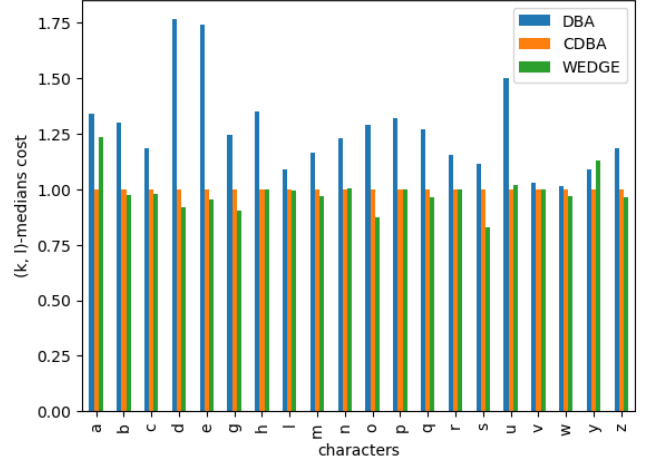


Figure 12: Comparison of improvement methods on handwritten character data set with $k = 2$, $\ell = 6$.

CDBA and DBA centers are more consistent with the data. In Figure 14 in the appendix, DBA obtains a significantly higher (k, ℓ) -medians cost than CDBA and the wedge method. The purple center trajectory computed by DBA seems “collapsed” compared to the centers computed by CDBA and the wedge method. These visual artifacts are similar to those shown in Figures 3 and 4.

We perform similar experiments for the pigeons data set and the storks migration data set. For both data sets we selected $k = 3$. For the pigeon data we used $\ell = 10$, while for the stork data we used $\ell = 14$. The (k, ℓ) -medians costs for FSA and DBA were significantly higher than those for CDBA and the wedge method, but figures for these were omitted due to space constraints. We focus on highlighting the visual artifacts in the clustering that we believe cause FSA and DBA to obtain these significant costs. In Figure 15 in the appendix, the artifacts of DBA are again clearly visible. The blue center produced by FSA seems to zigzag more than the centers produced by CDBA and the wedge method. The CDTW-based center improvement methods appear to give the smoothest and most natural blue center trajectories in these examples. In Figure 16 in the appendix, the blue centers produced by the CDBA and wedge methods are smoother and a better fit for the data than the center produced by DBA. The center trajectories produced by FSA bypass a sharp turn which appears to be a key feature of the data set. The CDBA and wedge methods nicely capture this feature.

Because of the visual artifacts in the characters, pigeons, and stork migration data sets, we are inclined to believe that CDBA and the wedge method produce superior cluster centers. For the same reason, we believe that the (k, ℓ) -medians cost is a reasonable measure to evaluate clustering algorithms, as it produces high costs on visually unintuitive cluster centers.

7 CONCLUSIONS

In this work we presented the first Continuous Dynamic Time Warping (CDTW) implementation that we know of. One reason why we are interested in a practical implementation of this distance

measure is enabling a center-based clustering approach that neither requires high-complexity representatives nor is sensitive to outliers. We conducted extensive experiments to evaluate our approach.

Our findings are that we should use the PAM method (partitioning around medoids) alongside the Gonzalez algorithm for initializing the centers. To compute an ℓ -simplification of the initial centers, the Imai–Iri approach using CDTW gives the best results, but the greedy approach using the Fréchet distance gives the best running time performance. We provided two novel methods of updating the initial cluster centers using the CDTW measure. Our CDTW-based update method avoids many of the artifacts that appear in DTW or Fréchet-based update methods, especially in the presence of outliers or low-complexity centers.

Interesting future work includes finding algorithmic approaches that lead to faster running times for CDTW computations and extending the usage of this distance measure to fields where DTW is currently predominantly used.

ACKNOWLEDGMENTS

The work of Aleksandr Popov is funded by the Dutch Research Council (NWO) under the project number 612.001.801.

REFERENCES

- [1] Pankaj K. Agarwal, Sarel Har-Peled, Nabil H. Mustafa, and Yusu Wang. 2005. Near-Linear Time Approximation Algorithms for Curve Simplification. *Algorithmica* 42, 3 (2005), 203–219. <https://doi.org/10.1007/s00453-005-1165-y>
- [2] Kevin Buchin, Anne Driemel, Joachim Gudmundsson, Michael Horton, Irina Kostitsyna, Maarten Löffler, and Martijn Struijs. 2019. Approximating (k, ℓ) -Center Clustering for Curves. In *30th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA2019)*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 2922–2938. <https://doi.org/10.1137/1.9781611975482.181>
- [3] Kevin Buchin, Anne Driemel, and Martijn Struijs. 2020. On the Hardness of Computing an Average Curve. In *17th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2020, Vol. 162)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, Article 19, 19 pages. <https://doi.org/10.4230/LIPIcs.SWAT.2020.19>
- [4] Kevin Buchin, Anne Driemel, Natasja van de L’Isle, and André Nusser. 2019. Kcluster: Center-Based Clustering of Trajectories. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL ’19)*. Association for Computing Machinery, New York, NY, USA, 496–499. <https://doi.org/10.1145/3347146.3359111>
- [5] Maike Buchin. 2007. *On the Computability of the Fréchet Distance Between Triangulated Surfaces*. Ph.D. Dissertation. Freie Universität Berlin. <https://doi.org/10.17169/refubium-6111>
- [6] Anne Driemel, Amer Krivošija, and Christian Sohler. 2016. Clustering Time Series Under the Fréchet Distance. In *27th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA2016)*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 766–785. <https://doi.org/10.1137/1.9781611974331.ch55>
- [7] Dheeru Dua and Casey Graff. 2017. *UCI Machine Learning Repository*. University of California, Irvine. Retrieved 2020-09-15 from <http://archive.ics.uci.edu/ml>
- [8] Alon Efrat, Quanfu Fan, and Suresh Venkatasubramanian. 2007. Curve Matching, Time Warping, and Light Fields: New Algorithms for Computing Similarity Between Curves. *Journal of Mathematical Imaging and Vision* 27, 3 (2007), 203–216. <https://doi.org/10.1007/s10851-006-0647-0>
- [9] Teofilo F. Gonzalez. 1985. Clustering to Minimize the Maximum Intercluster Distance. *Theoretical Computer Science* 38 (1985), 293–306. [https://doi.org/10.1016/0304-3975\(85\)90224-5](https://doi.org/10.1016/0304-3975(85)90224-5)
- [10] Sarel Har-Peled. 2011. *Geometric Approximation Algorithms*. Number 173 in Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, USA. <https://doi.org/10.1090/surv/173>
- [11] Hiroshi Imai and Masao Iri. 1988. Polygonal Approximations of a Curve—Formulations and Algorithms. In *Computational Morphology: A Computational Geometric Approach to the Analysis Of Form*, Godfried T. Toussaint (Ed.). Machine Intelligence and Pattern Recognition, Vol. 6. Elsevier Science, Amsterdam, Netherlands, 71–86. <https://doi.org/10.1016/B978-0-444-70467-2.50011-4>
- [12] Leonard Kaufman and Peter J. Rousseeuw. 1987. Clustering by Means of Medoids. In *Statistical Data Analysis: Based on the L_1 -Norm and Related Methods*. Elsevier Science, Amsterdam, Netherlands, 405–416.
- [13] Leonard Kaufman and Peter J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Hoboken, NJ, USA. <https://doi.org/10.1002/9780470316801>
- [14] Anil Maheshwari, Jörg-Rüdiger Sack, and Christian Scheffer. 2018. Approximating the Integral Fréchet Distance. *Computational Geometry* 70–71 (2018), 13–30. <https://doi.org/10.1016/j.comgeo.2018.01.001>
- [15] Richard Mann, Robin Freeman, Michael Osborne, Roman Garnett, Chris Armstrong, Jessica Meade, Dora Biro, Tim Guilford, and Stephen Roberts. 2010. Objectively Identifying Landmark Use and Predicting Flight Trajectories of the Homing Pigeon Using Gaussian Processes. *Journal of The Royal Society Interface* 8, 55 (2010), 210–219. <https://doi.org/10.1098/rsif.2010.0301>
- [16] Mario E. Munich and Pietro Perona. 1999. Continuous Dynamic Time Warping for Translation-Invariant Curve Alignment with Applications to Signature Verification. In *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV ’99, Vol. 1)*. IEEE, Piscataway, NJ, USA, 108–115. <https://doi.org/10.1109/ICCV.1999.791205>
- [17] OpenStreetMap contributors. 2020. Map data. Retrieved 2020-09-16 from <https://openstreetmap.org> Data available under ODbL 1.0.
- [18] François Petitjean, Alain Ketterlin, and Pierre Gançarski. 2011. A Global Averaging Method for Dynamic Time Warping, with Applications to Clustering. *Pattern Recognition* 44, 3 (2011), 678–693. <https://doi.org/10.1016/j.patcog.2010.09.013>
- [19] Chandan K. Reddy and Bhanukiran Vinzamuri. 2014. A Survey of Partitional and Hierarchical Clustering Algorithms. In *Data Clustering: Algorithms and Applications*, Charu C. Aggarwal and Chandan K. Reddy (Eds.). CRC Press, Boca Raton, FL, USA, Chapter 4, 87–110. <https://doi.org/10.1201/9781315373515>
- [20] Günter Rote. 2014. Lexicographic Fréchet Matchings. Presented at 30th European Workshop on Computational Geometry (EuroCG 2014). Retrieved 2020-06-23 from https://www.cs.bgu.ac.il/~eurocg14/papers/paper_56.pdf
- [21] Shay Rotics, Michael Kaatz, Sonda Turjeman, Damaris Zurell, Martin Wikelski, Nir Sapir, Ute Eggers, Wolfgang Fiedler, Florian Jeltsch, and Ran Nathan. 2018. Data from: Early Arrival at Breeding Grounds: Causes, Costs and a Trade-off with Overwintering Latitude. Movebank Data Repository. <https://doi.org/10.5441/001/1.v8d24552>
- [22] Shay Rotics, Michael Kaatz, Sonda Turjeman, Damaris Zurell, Martin Wikelski, Nir Sapir, Ute Eggers, Wolfgang Fiedler, Florian Jeltsch, and Ran Nathan. 2018. Early Arrival at Breeding Grounds: Causes, Costs and a Trade-off with Overwintering Latitude. *Journal of Animal Ecology* 87, 6 (2018), 1627–1638. <https://doi.org/10.1111/1365-2656.12898>
- [23] Erich Schubert and Peter J. Rousseeuw. 2019. Faster k -Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. In *Similarity Search and Applications (SISAP 2019)*. Springer, Heidelberg, Germany, 171–187. https://doi.org/10.1007/978-3-030-32047-8_16
- [24] Ben H. Williams. 2008. *Character Trajectories Dataset*. University of Edinburgh. Retrieved 2020-09-16 from <https://archive.ics.uci.edu/ml/datasets/Character+Trajectories>

A ADDITIONAL FIGURES

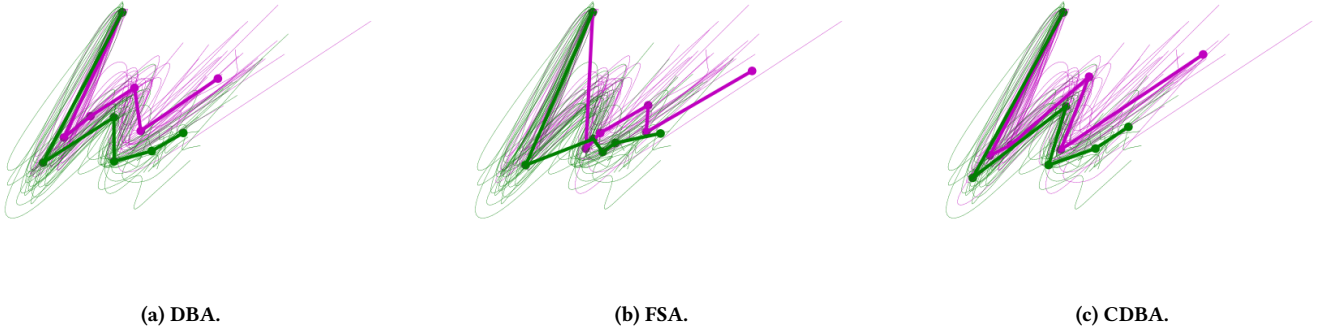


Figure 13: $(2, 6)$ -clusterings of h characters computed with DBA, FSA and CDBA.

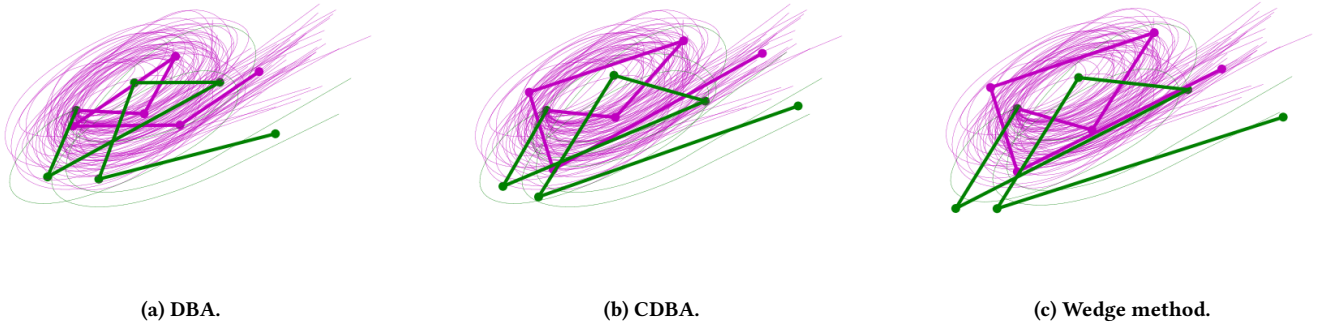


Figure 14: $(2, 6)$ -clusterings of e characters computed with DBA, CDBA and the wedge method.

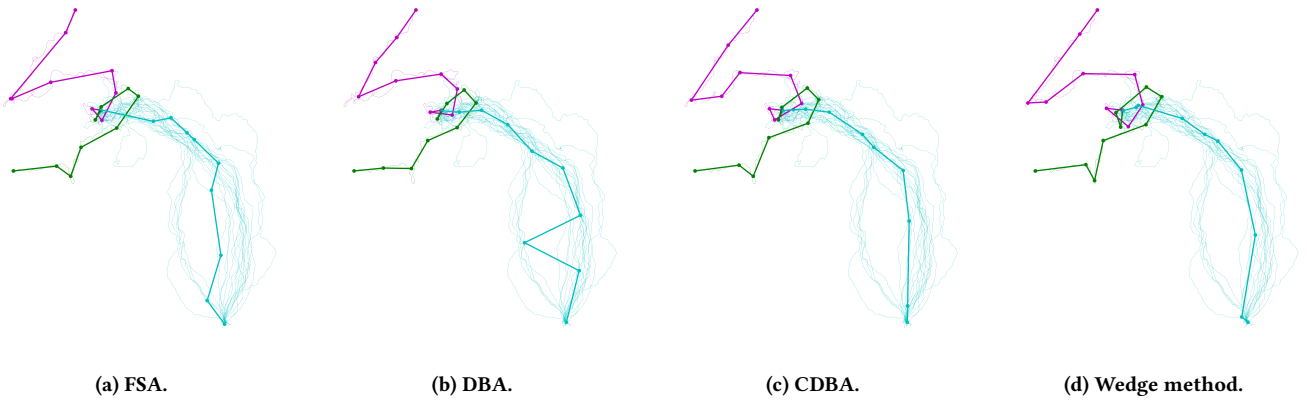


Figure 15: Centers of an initial clustering updated with (from left to right) FSA, DBA, CDBA, and the wedge method. The trajectory data used is the *brc* pigeon from the pigeon data set.

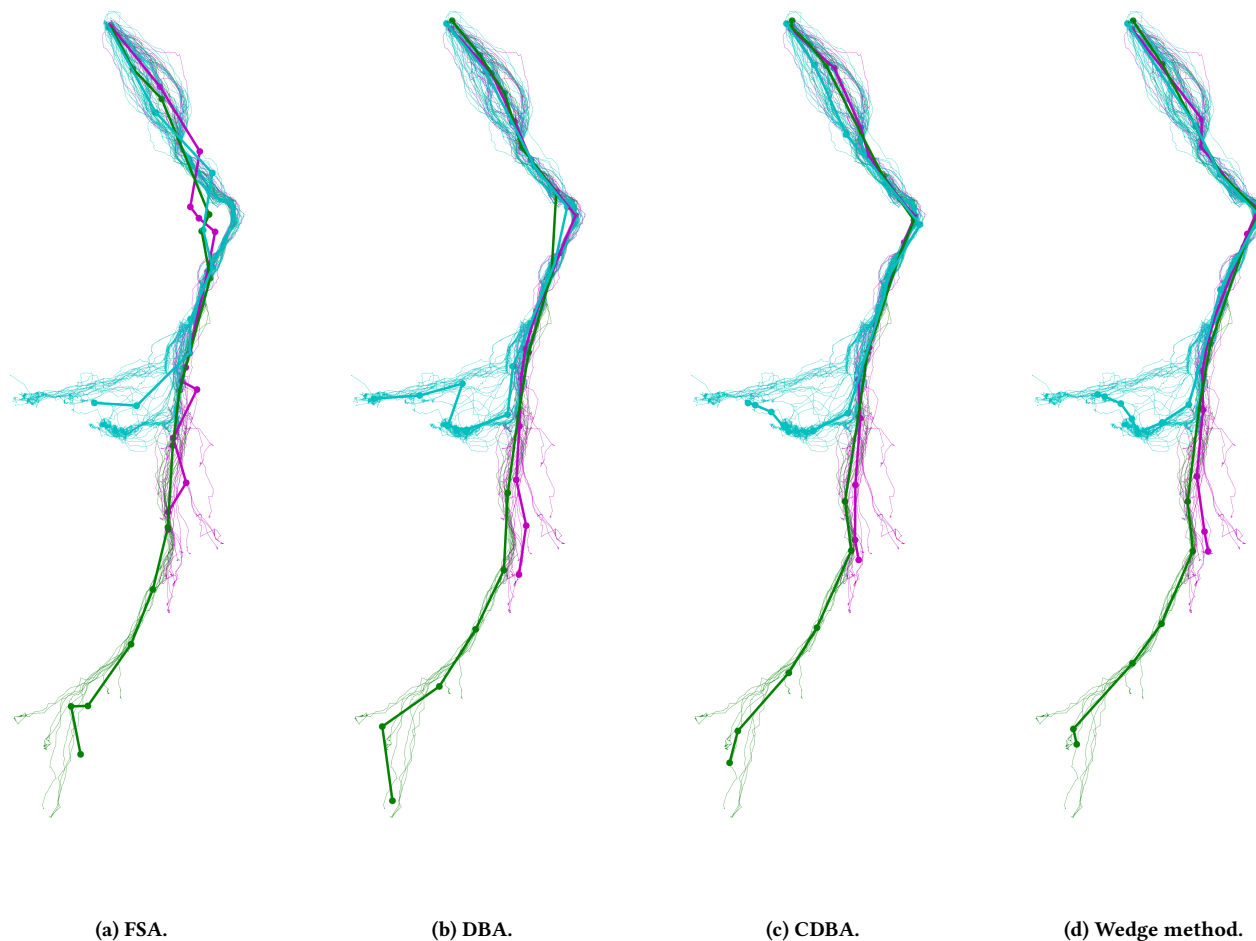


Figure 16: Centers of an initial clustering updated with (from left to right) FSA, DBA, CDBA, and the wedge method. The trajectory data used is the movebank stork migration data set.