TS-SCAN: a Density-Based Clustering Method for Trajectory Analysis and Anomaly Detection

Léo Monnier*
Thales DMS
Université Technologique de Troyes
Troyes, France
leo.monnier@thalesgroup.com

Louis-Marie Traonouez
Thales LAS La Ruche
Rennes, France
louis-marie.traonouez@thalesgroup.com

Karim Tit*
Inria Rennes
Rennes, France
karim.tit@inria.fr

***Abstract*—The surging amount of trajectories from sensors, antennas, and radars calls for the development of modern trajectory analysis techniques. In this context, we have introduced 2 ways to compare trajectories, which include a regularisation for dynamic time warping (DTW) and a faster computation of the Fréchet distance. As an example, we present TS-SCAN, an extension of the DBSCAN algorithm, which computes trajectory clusters and their average scheme in a database.**

***Index Terms*—clustering, trajectory analysis, anomaly detection, time-series metrics**

I. Introduction

Nowadays, sensors, radars, and antennas are everywhere and track a large number of moving objects, each identified by their trajectories. Worldwide traffic of airplanes and ships, for instance, can be supervised via open data networks, using either the ADS-B (Barsheshat, 2011) or the AIS (Balduzzi et al., 2014) protocols. The exploitation of these massive datasets opens the way for numerous applications, from trajectory optimization, to automatic classification and identification, or anomaly detection. For this reason, this topic is of interest to both civil and defense activities.

The analysis of trajectories requires specific data mining techniques adapted to process these composite objects, each composed of a variable number of detections (Wang et al., 2020). Specifically, metrics from time series, such as Dynamic Time Warping (Müller, 2007) and Discrete Fréchet distance (Eiter and Mannila, 1994), are particularly suitable for comparing trajectories. With adequate metrics, it is then possible to use various classical machine learning algorithms to analyse trajectories.

The idea behind this work is to propose a way to compare trajectories efficiently, first using an improved version of the well-known DTW metric, with a regularization term that makes it length invariant ($aDTW$). Alternatively, we propose a faster computed discrete Fréchet distance (fastdfrechet) that works with fewer assumptions than other found optimizations. We show that these two metrics work well and are complementary for the comparison of multi-sensor trajectory data, particularly when applied to aircraft trajectories.

*The majority of the work was carried out during the authors' stay at the "La Ruche" center of Thales LAS.

Then, using these two metrics, we propose a density-based clustering algorithm for trajectories. The idea behind the algorithm is to exploit the divergence matrix between trajectories, computed with $aDTW$ or fastdfrechet, and then to compute the representation of trajectory clusters with an approach based on Dynamic Based Averaging (DBA). As an example, we applied this data-mining technique to an ADS-B database to obtain the general scheme of trajectories.

This paper is organized as follows. In the first part, we present related work on the subject. Then, we give the technical background on metrics, followed by an explanation of our contributions regarding the two chosen metrics, and we introduce, in the fifth part, an adapted density-based clustering algorithm for trajectory analysis. Afterwards, we present an application of this algorithm on the ADS-B trajectories database and test anomaly detection with the chosen metrics. We conclude with the perspectives for antennas and radar surveillance.

II. Related Work

As previously stated, with the rise of open data touching many domains (aircraft, ships, buses, taxis...), a plethora of metrics, clustering algorithms, and anomaly detection frameworks dedicated to trajectories have emerged.

A recent survey from Bian et al. (2018) compares a set of metrics for trajectory analysis, including DTW or Discrete Fréchet distance, and it exposes accurate clustering methods such as DBSCAN. Another survey from Tao et al. (2021) is dedicated to metrics for trajectory comparison. Again, in the work of P. C. Besse et al. (2016), different metrics are compared, and clustering algorithms are applied to a taxi database.

About the two metrics used in our paper, the works of Salvador and Chan (2007) and Etienne et al. (2017) also proposed algorithmic optimization by constraining alignment exploration, similar to our approach. The algorithm for Discrete Fréchet Distance is based on the assumption that the two trajectories are already well aligned. Our method fastdfrechet allows us to relax this assumption.

Moreover, to validate our approach we used the application of clustering with DTW in the work of Petitjean et al. (2011) or with k-means algorithm in the work of Ma et al. (2023).

In air traffic trajectory analysis and clustering, we refer the reader to the papers of Olive and Morio (2018) and Basora et al. (2017), in which fully constructed frameworks using HDBSCAN are presented. The majority of these papers find that DBSCAN is well-suited for trajectory clustering.

III. Technical Background for Metrics

Trajectories in this paper are seen as sequences of coordinates and we chose to apply two metrics for sequence comparison. These metrics only focus on movement and do not detect time delay or velocity difference. They are well-known metrics for time series analysis and seem to apply well to multi-sensor trajectories, which are indeed specific types of time series.

A. Dynamic Time Warping

Dynamic Time Warping, also known as DTW, is a metric that searches the optimal alignment between all elements of two trajectories and sums the distance between them. It is defined as follow for two time series $X = \{x_1, ..., x_n\}$ and $Y = \{y_1, ..., y_m\}$ and $d()$ a distance function between two points:

$$DTW(X,Y) = \min_{\pi \in \mathcal{A}(X,Y)} \sum_{(i,j)\in\pi} d(x_i, y_j)$$

with $\mathcal{A}(X,Y)$ the set of all admissible paths, also called *warping* paths, between X and Y. In the scope of this paper, a warping path is a finite sequence of index pairs, denoted by $\pi = ((i_0, j_0), (i_1, j_1), \ldots, (i_K, j_K))$, satisfying the following properties:

1) $i_0 = j_0 = 1$ and $i_K = n, j_K = m$
2) For all $k \in \{1, \ldots, K\}$, $(i_k, j_k) = (i_{k-1} + 1, j_{k-1})$ OR $(i_k, j_k) = (i_{k-1}, j_{k-1} + 1)$ OR $(i_k, j_k) = (i_{k-1} + 1, j_{k-1} + 1)$.

This metric is well suited to compare trajectories of similar length, but otherwise, it becomes inadequate.

B. Discrete Fréchet distance

The Fréchet Distance is a metric commonly used for measuring divergence between functions. The discrete variant of this metric, as introduced in the work of Wien et al. (1994), can be used to detect the punctual differences between two sequences. It is defined as follows:

$$dFrechet(X,Y) = \min_{\pi\in\mathcal{A}(X,Y)} \max_{(i,j)\in\pi} d(x_i, y_i)$$

with $\mathcal{A}(X,Y)$ defined in the same way as seen in the definition of DTW. In contrast to DTW, $dFrechet$ works well despite variations in trajectory length in the database. The limitation arises in our case study on trajectories due to the requirement for numerous comparisons.

IV. Contributions on Metrics

In this section, we analyzed the effectiveness of the two previously defined metrics for the comparison of trajectories and proposed two main adaptations of these metrics to improve their applicability on a database of heterogeneous trajectories. In every data analysis context, it is interesting to have metrics that represent the mean or maximum distance between elements. In this idea, we can easily imagine that DTW could be enhanced to represent the mean and $dFrechet$ the maximum distance in this context.

A. Regularization for Dynamic Time Warping

Analyzing the formula of DTW, we notice that the value of DTW depends on the length of π, and it follows that this metric is dependent on the two compared trajectories' lengths. It means that the comparison between the two values of this metric is limited. This dependency is a critical point in the comparison of trajectories from multi-sensor types detections or multi-object trajectories. To limit this impact we introduced a regularization term in the formula of DTW, defined as follows:

$$\pi^*(X,Y) = \operatorname*{arg\,min}_{\pi\in\mathcal{A}(X,Y)} \sum_{(i,j)\in\pi} d(x_i, y_i)$$

$$aDTW(X,Y) = \frac{1}{|\pi^*(X,Y)|} \sum_{(i,j)\in\pi^*(X,Y)} d(x_i, y_i)$$

where $|\pi^*(X,Y)|$ denotes the length of the path $\pi^*(X,Y)$. Fundamentally, $aDTW$ represents the minimal distance between all alignments of the two compared trajectories, scaled by the optimal alignment path size. From a computational perspective, the search of the optimal path $\pi^*(X,Y)$ and of the estimation of $aDTW(X,Y)$ can be performed simultaneously, since many DTW implementations, like the FastDTW method introduced by Salvador and Chan (2007), return both the value and the alignment path.
Another argument to prefer $aDTW$ in place of DTW is that the value returned by $aDTW$ is in the same order of magnitude as the distance between points on a map.

B. Discrete Fréchet Distance Optimisation

As seen previously, having a metric that detects punctual divergence between trajectories is important for anomaly detection, but many implementations of Discrete Frechet Distance are computationally costly. Based on the work made on FastDTW, we propose a version of FastdFrechet, presented in Algorithm 1, which is of $\mathcal{O}(n)$ in time and space complexity as FastDTW is, instead of $\mathcal{O}(n^2)$. This method is not based on the idea that the two compared trajectories are already well aligned but hopes that the alignment is close to the one obtained on a subscale of the two trajectories. It returns an approximation of $dFrechet$, controlled by radius parameter r.

The idea behind that is to limit the number of alignments evaluated. To do so, we compute $dFrechet$ on the lower resolution of the two compared trajectories and we search the solution in a restricted environment depending on the radius

Algorithm 1 $FastDFrechet$

Input X, Y two-time series, r a distance to search for from a warping path of a sub-resolution $DFrechet$

Output $Dist$, π_F^* the optimal warping path

$MinTSize = radius + 2$
$Window := \{(x_i, y_j)|x_i \in X, y_j \in Y\}$
if $len(X) \geq MinTSize$ **and** $len(Y) \geq MinTSize$ **then**
  $X_{shrunked} = downscale(X)$
  $Y_{shrunked} = downscale(Y)$
  _ , π_F^*= $FastDFrechet(X_{shrunked}, Y_{shrunked}, r)$
  $Window := \{(x_{i\pm k}, y_{j\pm l})|(i,j) \in \pi_F^*, k,l \in [0,r]^2\}$
end if
$Dist, \pi_F^* = dFrechet(X, Y, Window)$

parameter for a better resolution of these trajectories. With this computation, we hope that the punctual divergence between trajectories will be detected in their lower resolution.

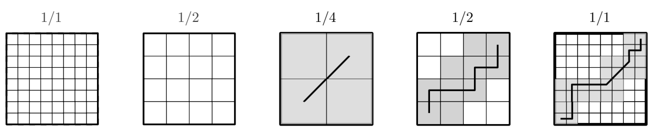

Fig. 1. Evolution of warping path for FastDfrechet for radius 1

For better visualization, Fig1 shows the evolution of the exploration domain for alignment during the algorithm for an arbitrary example.

V. DBSCAN FOR TRAJECTORIES

In this part, we introduce a density-based clustering for trajectories analysis. Density-based clustering is a well-known data clustering algorithm. We chose to use DBSCAN, instead of for instance K-means, another well-known clustering algorithm, because in the trajectories database, the number of clusters is generally not known. On the other hand, the two parameters required by the algorithm, namely the maximum distance between two neighbors' trajectories, and the minimum number of trajectories in a cluster, can be easily determined. Applied to trajectories analysis the DBSCAN algorithm detects recurring trajectories and represents their general path or raises abnormal or punctual trajectories.

A. Divergence matrix computation

In the distance matrix computation, we have the choice between the two previous metrics: Fréchet distance or $aDTW$. $aDTW$ focuses more on the general movement of trajectories because it represents the average distance between trajectories, while small variations will be hidden. On the contrary, Fréchet distance focuses more on punctual divergence between trajectories and will detect any little detour. We recommend the use of $aDTW$ in a general analysis of the database, for example, a search of general movement between airports for our case study, and Fréchet distance for anomaly detection.

B. Application of DBSCAN on trajectories

The idea of DBSCAN for trajectory analysis is to generate groups of similar movements in the database. It detects two types of elements: recurrent routes represented by clusters, and specific routes traveled only a few times and detected as noise.
To do so, it takes two information in input: the maximum distance for two trajectories to be called close, and the number of close trajectories to initiate a cluster. It follows these steps for each trajectory X :

1) Detect the set $T_d(X)$ of trajectories at a distance d of X.
2) If in $T_d(X)$ there is an element which is in a cluster $\mathcal{C}_i$, X is added to the cluster. If elements from more than one cluster appear in $T_d(X)$, all the clusters are merged together.
3) Else if there are k elements in T_d and none of them appear in a cluster, it creates a cluster $\mathcal{C}_i$ with all these elements.
4) Else, X is put in the noisy trajectories group.

At the end, the noisy trajectories group represents the non-recurrent trajectories defined by the criteria given and the clusters represent each recurrent pathway on the database.

C. Clusters representation

Since every trajectory in a cluster may have a different length, we cannot represent the cluster by its centroïd. Instead, a common technique to represent clusters of complex objects is to determine the medoïd of the cluster, which is the closest element to all other elements in the cluster.

One of the newly used methods to average time series clusters based on DTW is Dynamic Barycenter Averaging (DBA) from the work of Petitjean et al. (2011). The idea is to repeat a certain amount of these steps

1) Take a central element c_i for the cluster C_i
2) Apply DTW between this element and trajectories from the cluster X_k to obtain the alignment path $\pi^*(c_i, X_k)$ defined in part 3.
3) For each coordinates of c_i make the barycenter of each coordinates of any X_k aligned with $\mathcal{C}_i$ with all this elements.

Our approach is to initialize the cluster with the medoïd of the cluster and make only one step of the algorithm. It seems to give a pretty good cluster representation and to limit the computation cost of the algorithm. The advantage of this method is that the cluster representation is based on each trajectory of the cluster instead of only one trajectory.

Figure 2 shows an example of the different cluster and their representation in black after using TS-Scan on a database of 69 elements from TSlearn (Tavenard et al., 2020).

VI. APPLICATION TO AIRSPACE SURVEILLANCE

To demonstrate the usefulness of our approach for trajectory analysis we applied these methods to an ADS-B database from (Olive and Basora, 2019) of 3551 aircraft trajectories of 50 coordinates each from Paris Orly airport to Toulouse airport.

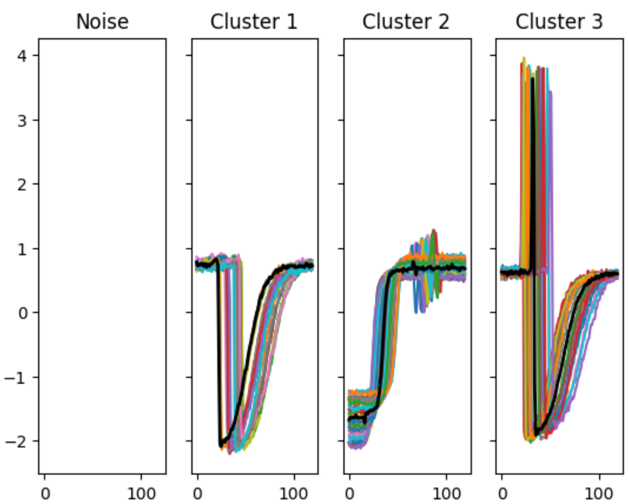

Fig. 2. Example on TSlearn database with $d = 0.05$ and $min_{elements} = 2$

A. Applying TS-SCAN

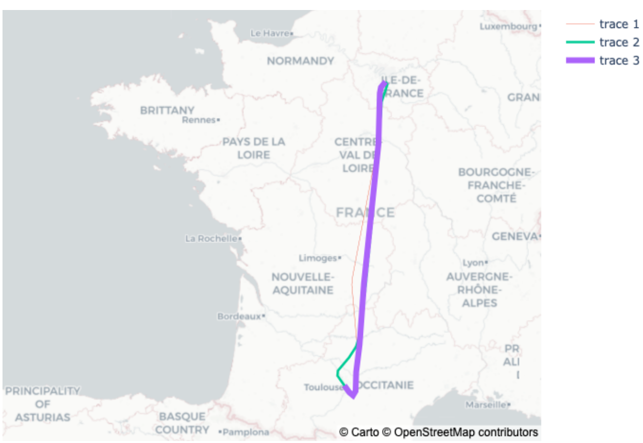

Fig. 3. Example of TS-SCAN with $d = 0.05$ and $min_{elements} = 10$ on a sample of ADS-B Trajectories.

Our first objective is to analyze the different routes for the commercial line and identify the different flight patterns for each route.

To do so, we apply TS-Scan with the metrics $aDTW$. The use of the chosen metrics is important for the interpretability of the DBSCAN parameter defined by the maximum distance for the direct association of trajectories. In our example, $d = 0.05$, and the minimum number of trajectories in a cluster is 10.

We obtain for each route several clusters of flight trajectories and we can get a representation of the clusters by computing the average trajectory using the method presented in section V-C. The width of each trajectory represents the number of flights using this road. We found reference trajectories for this commercial road. Moreover, the number of non-associated trajectories is 221.

Each cluster represents a different flight pattern, depending for instance on the runaway used and the runaway direction, and on the waypoints that each flight is assigned to follow. In this sense, the cluster representation is a way to identify the flight plans followed by the aircraft in question.

In this manner, the maximum distance parameter of TS-Scan strongly influences the patterns that may be observed. Using a small distance we can identify small deviations due to take-off and landing patterns, weather, or flight habits. Using a larger distance we focus mainly on big trajectory variations which allows for instance to identify flight routes above an area.

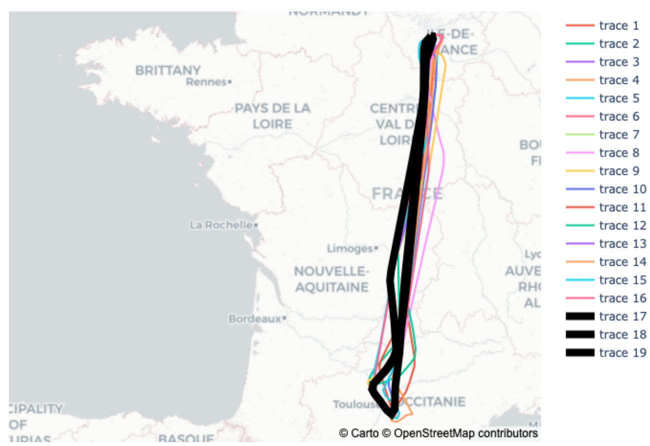

Fig. 4. TSscan on noise data with $d = 0.05$ and $min_{elements} = 2$

To see the distance of the noise data from reference trajectories we refer to Figure 4. Bigger width trajectories represent reference tracks and smaller represent non-detected clusters with the first use of TS-Scan from Figure 3.

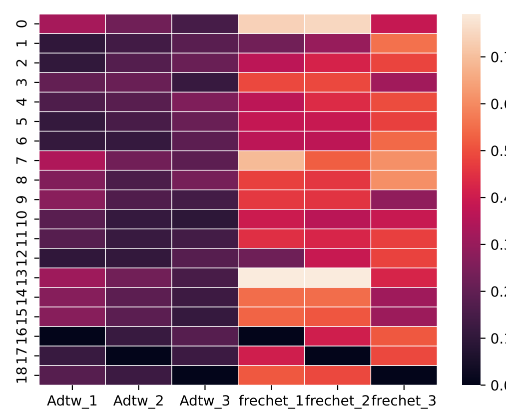

Fig. 5. Comparison of $FastDfrechet$ and $aDTW$ distances, between reference and real trajectories.

B. Trajectory deviation detection

For trajectory deviation detection, we recommend using $FastDFrechet$ to detect little variation between a trajectory and its reference. In our case study, we created an anomaly detection algorithm for trajectories that leverages an alert if the $FastDFrechet$ between the aircraft trajectory and its flight-plan value is over a limit s. We developed another version that compares the aircraft trajectory with a pool of averaged trajectories from the TS-SCAN algorithm with the same conditions as the flight plan. The main advantage of alerts raised by this algorithm is that they are fully explainable whereas many anomaly detection algorithms are not.

In Figure 5 we represent the computed $FastDFrechet$ and $aDTW$ distances between the 3 reference trajectories shown in Figure 3 (columns) and 19 actual trajectories shown in Figure 4 (lines). We can see that the Discrete Fréchet distances between reference trajectories and noisy trajectories are generally higher than the $aDTW$ distances, and, crucially, exhibit a wider and more discriminative range of variation, which is beneficial when setting thresholds for anomaly detection.

VII. Conclusion

Our proposed metrics seem promising in trajectory anomaly detection, and they are particularly suited for air traffic analysis. It allows for instance to automatically detect part of a trajectory that deviates from a flight plan or a reference track. The results of those metrics have the advantage of being explainable and easily understandable as their value is similar to the distance between two points in each trajectory.

With the proposed DBSCAN implementation, we further give a tool to analyze large trajectory databases. In airspace surveillance, it can be used to identify generic patterns in aircraft movements and detect outlier tracks on which operators could focus their time and attention.

For future work, we note that recent and more sophisticated variants of clustering algorithms, like the HDBSCAN, have not been explored within this work. Yet, we think they could be well suited to further optimize our cluster detection process.

References

Marco Balduzzi, Alessandro Pasta, and Kyle Wilhoit. A security evaluation of ais automated identification system. In *Proceedings of the 30th Annual Computer Security Applications Conference*, ACSAC '14, page 436–445, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450330053. doi: 10.1145/2664243.2664257.

Abraham A. Barsheshat. Implementation of ads-b systems — benefits and considerations. In *2011 Tyrrhenian International Workshop on Digital Communications - Enhanced Surveillance of Aircraft and Vehicles*, pages 197–201, 2011.

Luis Basora, Jérôme Morio, and Corentin Mailhot. A trajectory clustering framework to analyse air traffic flows. In *SID 2017, 7th SESAR Innovation Days*, 2017.

Jiang Bian, Dayong Tian, Yuanyan Tang, and Dacheng Tao. A survey on trajectory clustering analysis. *CoRR*, abs/1802.06971, 2018.

Thomas Eiter and Heikki Mannila. Computing discrete fréchet distance. 1994.

Laurent Etienne, Maxence Esnault, and Florian Lardy. Optimized discrete fréchet distance between trajectories. pages 11–19, 11 2017. doi: 10.1145/3150919.3150924.

Dongfang Ma, Bo Fang, Weihao Ma, Xiaodong Wu, and Sheng Jin. Potential routes extraction for urban customized bus based on vehicle trajectory clustering. *IEEE Transactions on Intelligent Transportation Systems*, 24(11):11878–11888, 2023. doi: 10.1109/TITS.2023.3288030.

Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

Xavier Olive and Luis Basora. Reference data sets for detection and identification of significant events in historical aircraft trajectory data. 12 2019. doi: 10.6084/m9.figshare.11406735.v1.

Xavier Olive and Jérôme Morio. Trajectory clustering of air traffic flows around airports. *Aerospace Science and Technology*, 84:776–781, 2018.

P. C. Besse, B. Guillouet, J. -M. Loubes, and F. Royer. Review and Perspective for Distance-Based Clustering of Vehicle Trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 17(11):3306–3317, November 2016. ISSN 1558-0016. doi: 10.1109/TITS.2016.2547641.

François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, 2011.

Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11(5): 561–580, oct 2007. ISSN 1088-467X.

Yaguang Tao, Alan Both, Rodrigo I. Silveira, Kevin Buchin, Stef Sijben, Ross S. Purves, Patrick Laube, Dongliang Peng, Kevin Toohey, and Matt Duckham. A comparative analysis of trajectory similarity measures. *GIScience & Remote Sensing*, 58(5):643–669, July 2021. ISSN 1548-1603. doi: 10.1080/15481603.2021.1908927.

Romain Tavenard, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, and Eli Woods. Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, 21(118):1–6, 2020.

Di Wang, Tomio Miwa, and Takayuki Morikawa. Big trajectory data mining: A survey of methods, applications, and services. *Sensors*, 20(16), 2020. ISSN 1424-8220. doi: 10.3390/s20164571.

Technische Universität Wien, Thomas Eiter, Thomas Eiter, Heikki Mannila, and Heikki Mannila. Computing discrete fréchet distance. Technical report, 1994.