Creating Air Corridors with Density Based Clustering Method

Oguzhan Ciftci[1] and Suat Ozdemir[2]

[1] Department of Command Control Software Design, Aselsan, Ankara, Turkey
[2] Computer Engineering Department, Gazi University, Ankara, Turkey

ARTICLE INFO	ABSTRACT
Keywords: *clustering, air corridors,* *dbscan, hdbscan, optics*	The danger of shooting civil aircrafts with artillery units is increasing day by day due to the number of war and skirmish zones in the world. Considering the danger of shooting a civilian aircraft, a safe flight zone must be defined for civilian aircrafts. Data mining techniques can quickly and accurately identify safe flight zones. In this study, we firstly discover the civilian aircraft busy routes using a novel dataset and several clustering algorithms. We perform a comparison among algorithms based on their ability to create air corridors.

1.   Introduction

The danger of shooting civil aircrafts with artillery units is increasing day by day due to the number of war and skirmish zones in the world. Considering the danger of shooting a civilian aircraft, the Turkish Armed Forces and NATO members must define a safe flight zone for civilian aircrafts. A safe flight zone is a rectangular prism, called "Air Corridor" that is obtained via a Command Control software. Air Corridor is defined as the route that civilian aircraft generally pass. A ballistic missile trajectory can be calculated without fire weapons. In the event of this calculation, a violation is provided in the form of a conflict between the ballistic missile trajectory and the Air Corridor. Thus, the civilian aircraft are allowed to pass through these dangerous battlefields.

As can be seen, air corridors are needed to be created in command and control software, given the increasing number of civil aircraft and the cost of creating air corridors during the war. The creation of air corridors will be possible by explore route patterns through machine learning methods from civil aircraft data collected from previous times. We use several unsupervised learning algorithms to solve this problem. DBSCAN algorithm give us a better solution than others. We trained all of our novel dataset with ~1.3 records and as a result 212 clusters obtained.

The rest of the paper is organized as follows: Section 2 introduces clustering algorithms, dataset and defines air corridors in detail. Section 3 present the experimental results of the clustering

_____

⬚ Corresponding Author E-Mail Address: ogciftci@aselsan.com.tr

algorithms and effect of feature addition. Finally, we conclude our work with demonstration of obtained air corridors in Section 4.

2. Preliminaries

a. Clustering Algorithms

Clustering can be considered the most important unsupervised learning problem. Clustering analysis try to find a structure in the unlabeled data collection. In this study, Density-based clustering algorithms used. Because we want to identify airplane location points where densities are higher in our dataset and create air corridors.
Density-based clustering algorithms DBSCAN (Density Based Spatial Clustering of Applications with Noise), OPTICS (Ordering Points To Identify Clustering Structure) and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithms are used.

DBSCAN Algorithm: The DBSCAN algorithm was presented by Ester, Kriegel, Sander and Xu at the KDD'96 conference [1]. This algorithm calculates clustering by grouping the objects with more objects than the predetermined threshold in a given region by calculating the distances of the objects with their neighbors. The DBSCAN algorithm has introduced many new approaches to data mining.

OPTICS Algorithm: The OPTICS algorithm was presented at the SIGMOD'99 conference by Ankerst, Breunig, Kriegel, and Sander [2]. It can be described as an improved version of the DBSCAN algorithm. In order to reduce the dependence on Eps and MinPts values which can be defined as the weakness of DBSCAN algorithm, it places data objects on a graph according to Eps value and finds clusters on the graph without needing MinPts value.

HDBSCAN Algorithm: The HDBSCAN is a hierarchical, density-based clustering algorithm that improves on previous density-based algorithms [3]. Its main output is a cluster hierarchy that describes the nested structure of density-based clusters in a dataset with respect to a single parameter, mpts.

b. Air Corridor

According to the Dictionary of the Military and Associated Terms Dictionary Air Corridor is that "A restricted air route of travel specified for use by friendly aircraft and established for the purpose of preventing friendly aircraft from being fired on by friendly forces". [4] Friendly forces like mortars, howitzers and MLRS (Multiple Launch Rocket System).

The Turkish Armed Forces, like every NATO member, have to define Air Corridors on the roads that these airplanes pass through, because all NATO members define Civil Aircraft as friendly. The purpose of creating air corridors is to check whether the ballistic missile trajectory formed by the artillery units, intersects this air corridor. If the trajectory intersects then Command Control Software will give violation. So, the shooting will not be allowed in this situation.

The Air Corridor is a rectangular prism that can be created with width and start-end locations in the minimum and maximum altitude range according to the altitude above sea level. An example is shown in figure 1.
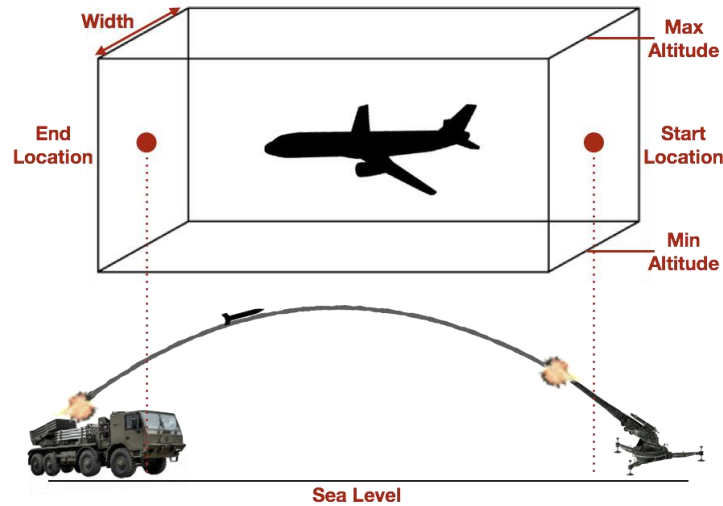
Fig. 1. Air Corridor definition with attributes.

However, the essence of our work is that the constitution of each Air Corridor to the Command Control Software, takes long times when the soldiers manually enter all the information of the Air Corridors, at present.

This method does not give good results considering the increasing number of Civil Aircraft and the cost of creating Air Corridor in case of war. In order to solve this problem, it is aimed to use a machine learning method to automatically create Air Corridors.

c. Description Of Dataset

Dataset is a 1-month data of civil aircraft flying over the Turkey above 30,000 feet. Turkey are in range: Latitude from 35.90 to 42.02 and longitude from 25.90 to 44.57 [5]. The dataset was collected at 1-minute intervals of Aircraft radar systems.

These systems give us very great information about aircrafts like speed, acceleration and location etc. So, we must decrease the size of dataset to trainable size on disk with data pre-processing.

The dataset with all attributes is ~3.6GB storage on disk as a JSON file format. With data pre-processing, we have 34 MB in size as uncompressed text, approximately ~1.3 million singular objects with 4 numerical attributes. The attributes are latitude, longitude, altitude and azimuth angle. The latitude and longitude values are in decimal degrees' format. All values in the dataset are min-max normalized before processing. The altitude value is in feet format and azimuth angle in degree format. In section X, we try to show our experimental observation about before and after adding azimuth angle information. The purpose of experiment is that adding extra attributes gets better result or worse result.

3. Experımental Analysis

In this section, we show some our experimental analysis about this study. At first, we compare clustering algorithms in order to output clusters. Secondly, we measure the performance of algorithms with certain measures. Finally, we have tried an experiment to look for the question of whether we can achieve better results.

Comparison of Clustering Algorithms Results

The input size of all the clustering algorithms we have compared in this study is 75000 airplane location point during flight and the minimum number of objects in eps radius value is 5. Since the eps parameter is not given in the hierarchical clustering methods, the DBSCAN algorithm eps value is given as 0.007. Outputs of the algorithms displayed in 3-dimension scatter plot with axis latitude, longitude and altitude normalized values. Figure 2 shows the output of the OPTICS algorithm (ε = 75000, minPts=5).
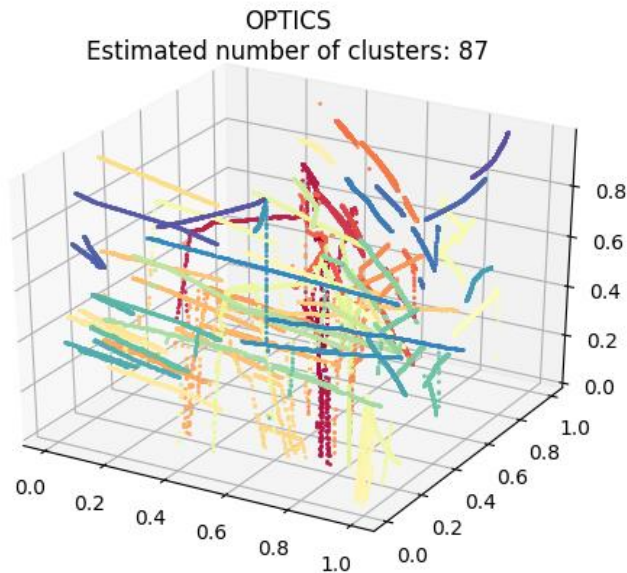

Fig. 2. OPTICS Algorithm output. Estimated number of clusters is 87.

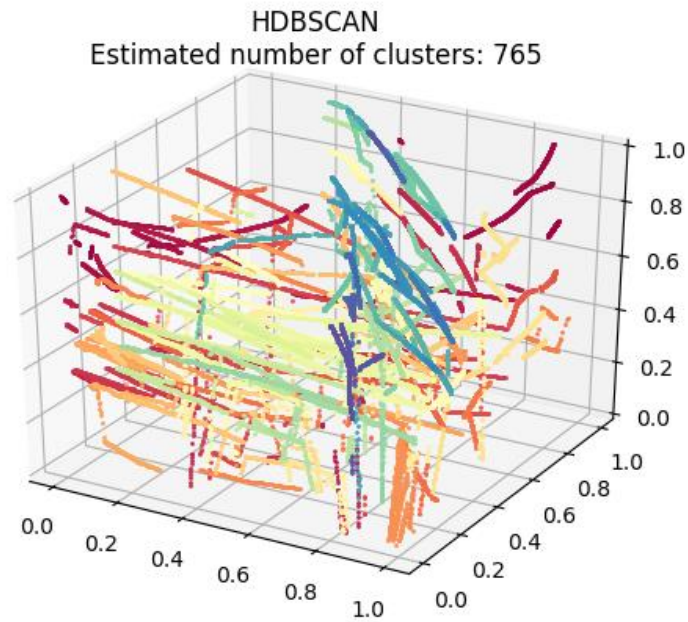Figure 3 shows the output of the HDBSCAN algorithm ($\varepsilon = 75000$, minPts=5).

Fig. 3. HDBSCAN Algorithm output. Estimated number of clusters is 765.

Figure 4 shows the output of the DBSCAN algorithm ($\varepsilon = 75000$, minPts=5, eps=0.007).
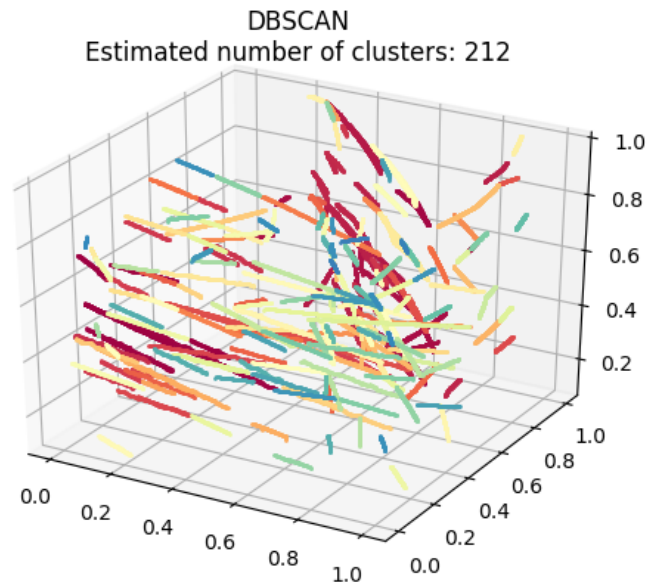

Fig. 4. DBSCAN Algorithm output. Estimated number of clusters is 212

As you can see in figure 2 and 3, hierarchical clustering methods are changing eps value to find clusters as hierarchical. So they are included in clusters at different altitudes points of the aircraft. However, we are trying to achieve simpler and more straightforward clusters of points that points at about the same altitude.

Thus, we can see from the DBSCAN algorithm, as seen, clusters of points that are flatter and at about the same altitude.
As a result, we decided to use the DBSCAN algorithm when air corridors were detected.

Performance Comparison

The total processing times of the algorithms have been analyzed. All experiments were run on a computer with 2 cores, 8 GB 1600 MHz DDR3 RAM and a 2.5 GHz processor.

These experiments were carried out using latitude, longitude and altitude which are features of the data set. Performance comparison table can be seen in Table 1.

The total processing times of the algorithms have been analyzed. All experiments were run on a computer with 2 cores, 8 GB 1600 MHz DDR3 RAM and a 2.5 GHz processor.
These experiments were carried out using latitude, longitude and altitude which are features of the data set. Performance comparison table can be seen in Table 1.

Input Size	OPTICS Cluster Count	Time	HDBSCAN Cluster Count	Time	DBSCAN Cluster Count	Time
5000	*49*	2 sec	51	0.11 sec	46	0.03 sec
10000	51	6 sec	67	0.22 sec	52	0.06 sec
25000	77	45 sec	263	0.83 sec	79	0.23 sec
50000	68	234 sec	908	1.84 sec	106	0.60 sec
75000	58	593 sec	1568	2.98 sec	121	1.12 sec
125000	*65*	2188 sec	3431	7.09 sec	160	2.41 sec

Tale 1.

A Feature Addition Experiment

At the beginning of our studies, we used latitude, longitude and altitude properties in our data set. However, expectation of our clusters to flatter, we observed that complex shapes became a single cluster. Since the air corridors are in the form of rectangular prisms, it is better for us to have the clusters in flat patterns.

To solve this problem, we added the azimuth angle feature of the aircraft to our dataset. At the output of our algorithm, we got flatter but more clusters. Because, this feature separates complex shaped clusters into flat clusters.

You can observe that a non-flat cluster in Fig. 5 is transformed into flat clusters by the addition of the azimuth angle feature.
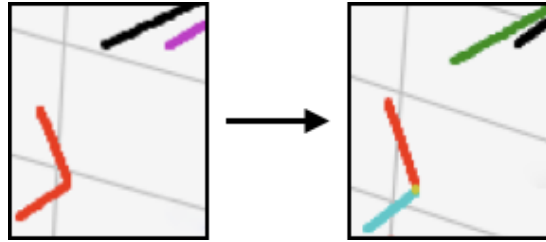

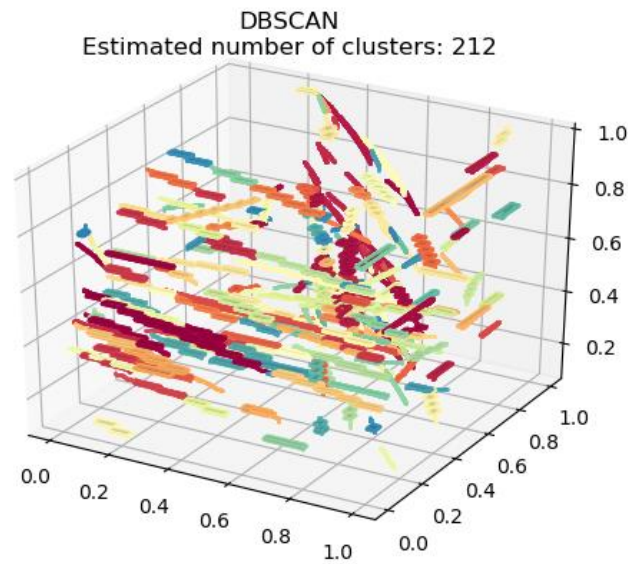**Fig. 5. Effect of addition azimuth angle feature on output.**

Fig. 6. Air Corridors of DBSCAN algorithm with parameters
($\varepsilon = 75000$, minPts=5, eps=0.007).

4. Result

In this study, we try to discover the civilian aircraft routes using a novel dataset and several clustering algorithms. We performed a comparison among algorithms based on their ability to create air corridors. We chose DBSCAN algorithm as it gives better results and runs faster than HDBSCAN and OPTICS algorithms. As a result, we have trained our algorithm with a learning dataset of ~1.3 million records. Using the output of our algorithm, we created air corridors. The clusters are framed with Air Corridors to use our Command and Control System. These air corridors

are presented in Figure 6. The results obtained in this study are expected to be used by military forces while they are practicing and/or in war cases.

References

[1]Ester, M., Kriegel, H., Sander, J. and Xu, X. (1996). A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp.226-231.

[2]Ankerst, Breunig, Kriegel, Sander, "OPTICS: Ordering Points to Identify the Clustering Structure", SIGMOD Conference, 1999.

[3]Campello, Ricardo JGB, Davoud Moulavi, and Jrg Sander." Density- based clustering based on hierarchical density estimates." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, 2013.

[4]DOD Dictionary of Military and Associated Terms. (2017). In: Air Corridor. p.9.

1. [5] Latitudelongitude.org. (2019). Turkey latitude and longitude. [online] Available at: http://latitudelongitude.org/tr/ [Accessed 19 Feb. 2019].