17th Hellenic Database Management Symposium (HDMS’19) 8-9 Jul 2019 @ Athens, Greece

Semantic-aware Aircraft Trajectory Prediction using Flight Plans

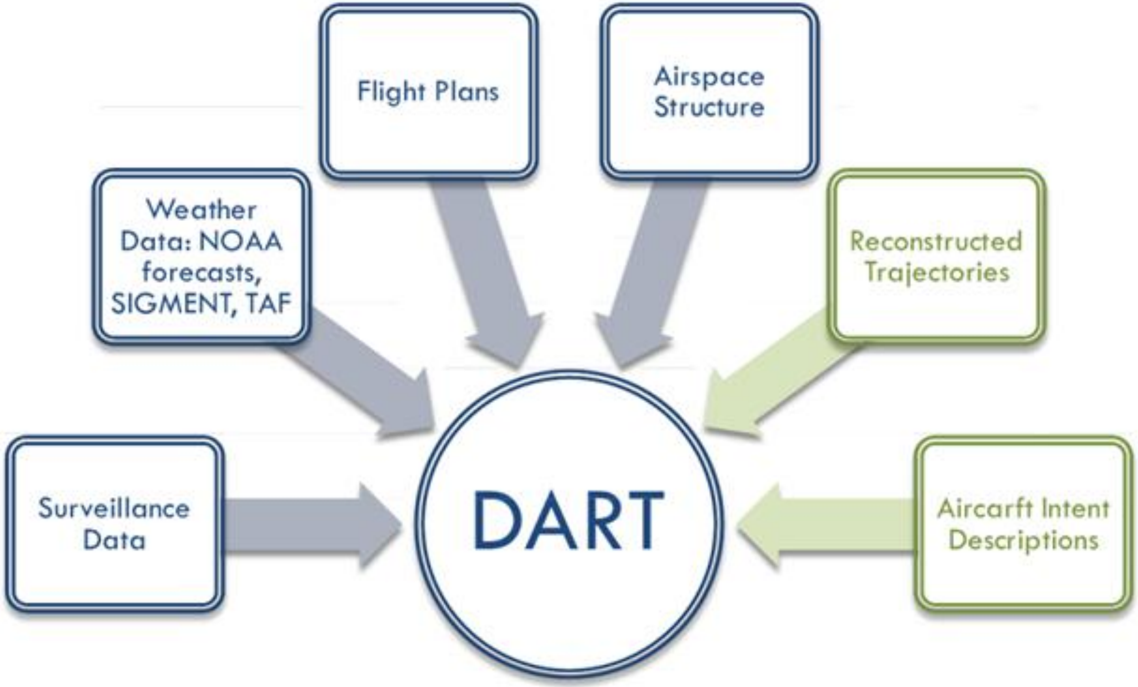
H. Georgiou, N. Pelekis, S. Sideridis, D. Scarlatti, Y. Theodoridis

Data Science Lab @ University of Piraeus, Greece

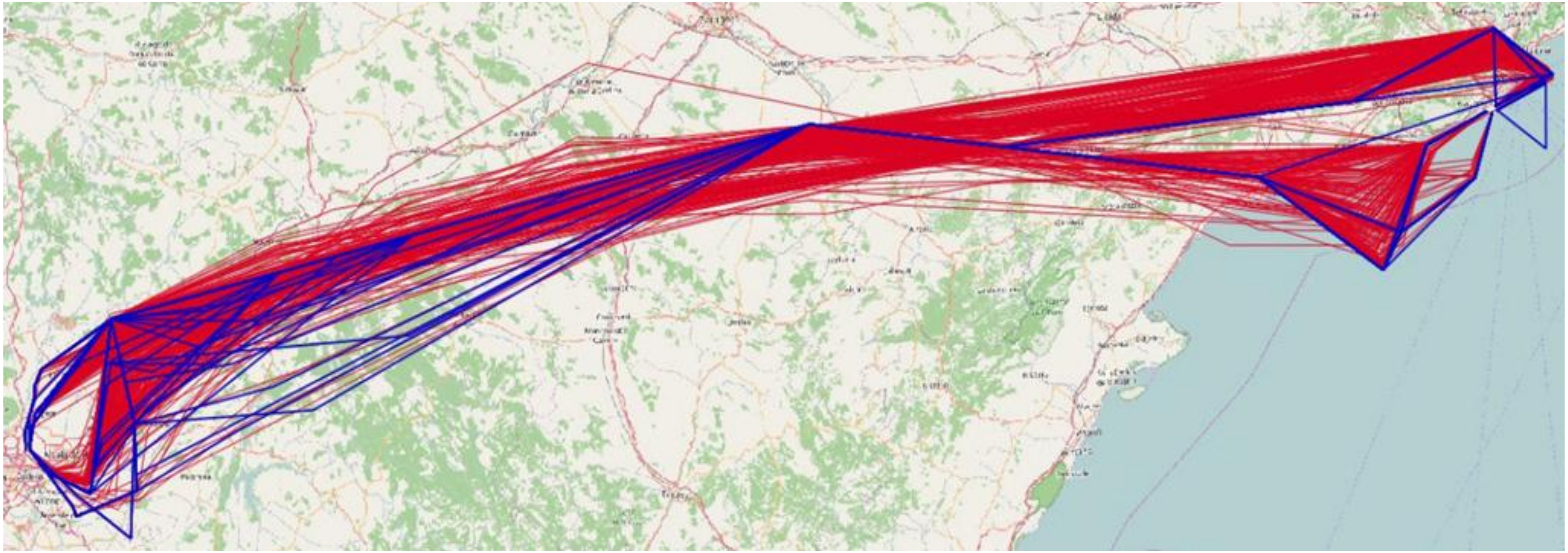
DATA STORIES

Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS RESEARCH CENTER

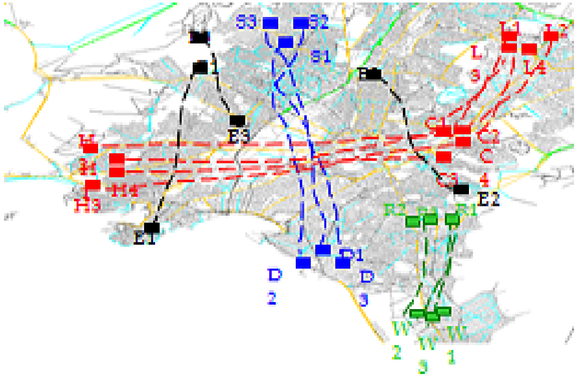


Example: Madrid/Barcelona route
(about 700 flights per month per direction)

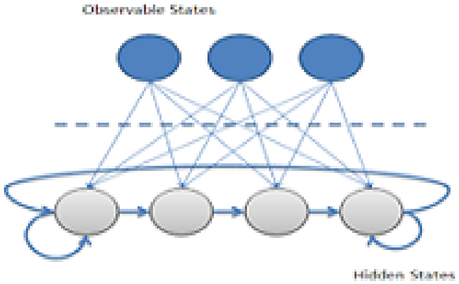


Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)

1st Step: Cluster semantically annotated trajectories



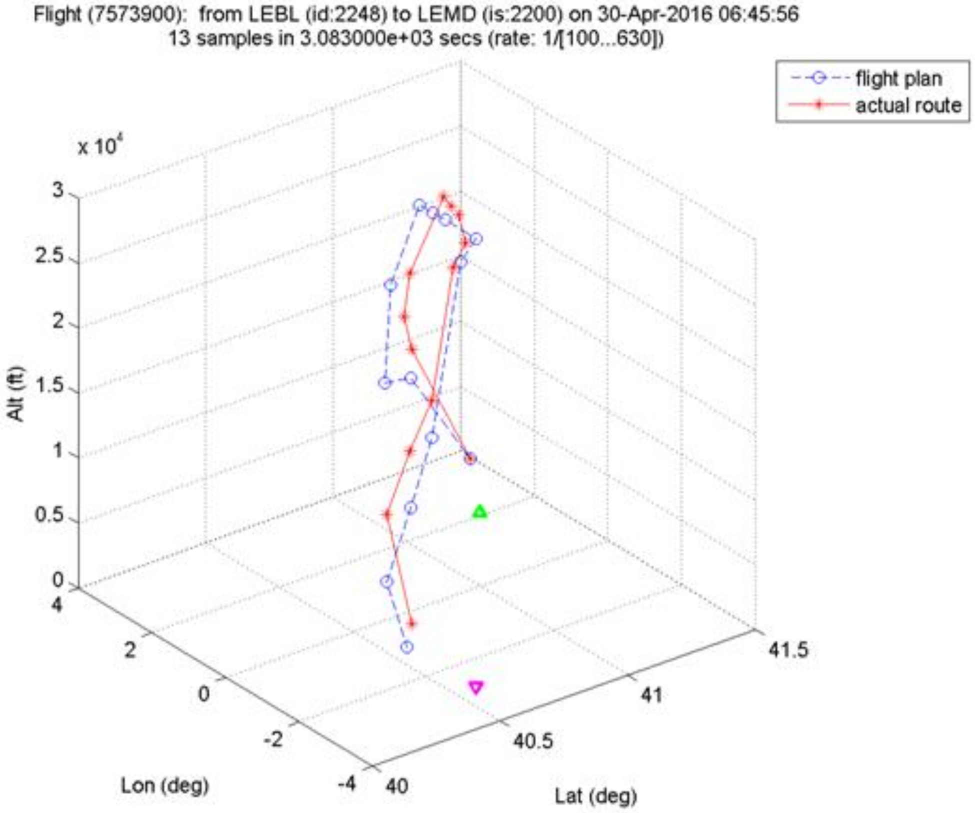
2nd Step: For each cluster, train a HMM



3rd Step: (Filter) Given a reference trajectory Q find top-k most probable HMM models

4th Step: (Refine) Similarity search among the trajectories that belong to the top-k HMMs

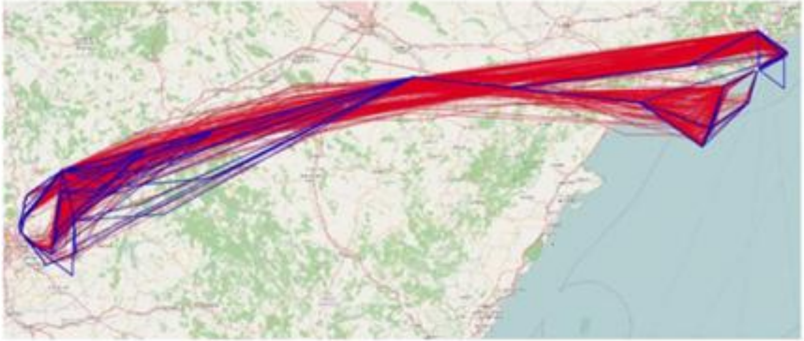
Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)



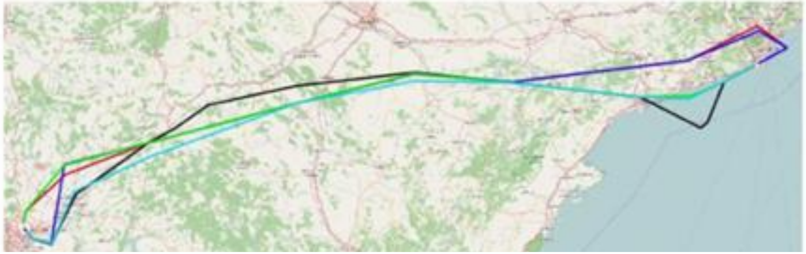
Basic idea – Method outline:

1. <u>Input</u>: Flight plans, actual routes, local weather, aircraft type, (...)
2. **Stage-1**: cluster the actual routes, produce semantic-aware medoids as "representatives"
3. **Stage-2**: Build a Pred.Model for each medoid, associate it with the cluster flight plans (emissions)
4. **Stage-3**: For a new flight plan, find the *k* closest matches (Pred.Model)
5. <u>Output</u>: $k \geq 1$ best-estim. (HMM: true, LR: synthetic) of the query FP, for prob.estim. or further refinement

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)



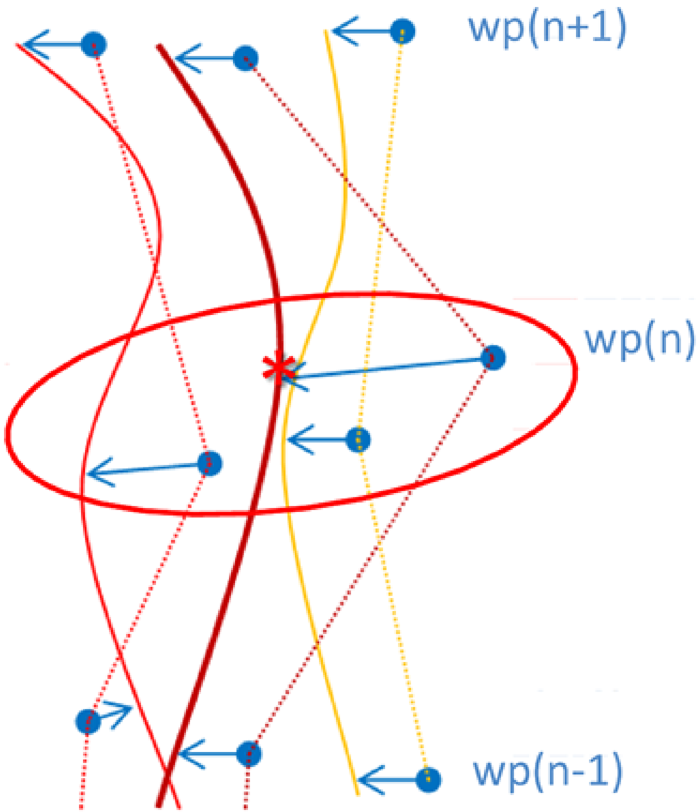
Input: flight plans (blue) and actual routes (red)



Step 1: clustering semantically annotated trajectories

- Input: Reference waypoints, matched with flight plans
- Each waypoint is enriched with semantics, i.e., local weather, aircraft type, etc.
- Use semantic-aware similarity metric, not just spatio-temporal track data
- Output: Semantic-aware cluster

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)



Stage-2, HMM approach:

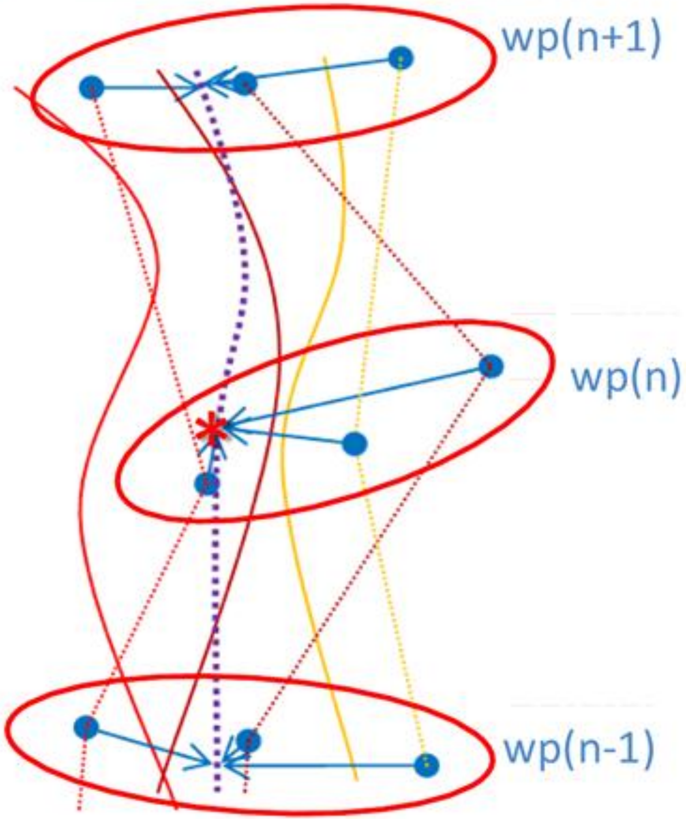
RT(n)-FP(n) statistics are used to build a probabilistic model (HMM emissions) for ref. point wp(n).

dotted line: flight plan (FP), solid line: actual route (RT)
arrows: FP/RT deviations, star: current pred. point (wp(n))
bold solid line: cluster medoid, pred. route for query FP

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)

DART

SESAR JOINT UNDERTAKING

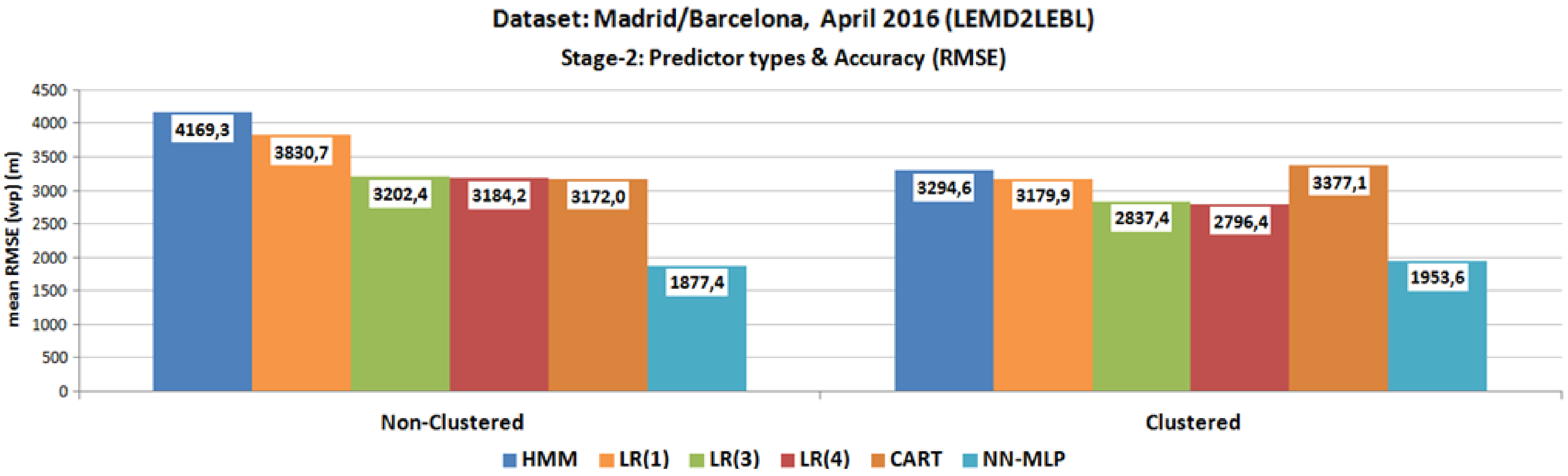


Stage-2, LR or DT approach:

RT(n) is estimated as synthetic from multiple/all FP(*) ref. points, used to build a LSE-minimum linear prediction model.

dotted line: flight plan (FP), solid line: actual route (RT)
arrows: FP/RT deviations, star: current pred. point (wp(n))
bold solid line: cluster medoid, pred. route for query FP

Overview: TP performance tests



- Input dimension: FP(1D) vs. full FP(1D) vs. full FP(3D) vs. full FP(3D)+AP:
  - LR/FP(3D) provides x3-x4 improvement of accuracy than LR/FP(1D) over HMM
  - LR/FP(4D), i.e. LR/FP(3D) + AP, improves accuracy only marginally (1-3%)
- Non-clustered vs. Clustered dataset (HMM, LR, CART):
  - Clustering improves accuracy across all models & dimensionality (14-26,5%)

Non-linear Regressors (NN)

DART

SESAR JOINT UNDERTAKING

			NN/MLP/h10(BR): RMSE (m)			
CLUSTER	Nk	Output:	RT(j)Lat	RT(j)Lon	RT(j)Alt	RT(j)R3D
(*)	696	Input:				
CVk	Nwp	wp1	240,7	139,3	31,5	**279,9**
(90/10)	11	wp2	105,0	71,1	67,1	**143,5**
		wp3	1500,4	301,2	221,6	**1546,3**
		wp4	2813,7	1424,6	244,0	**3163,2**
		wp5	1709,7	1298,6	342,8	**2174,2**
		wp6	1519,2	670,9	235,9	**1677,4**
		wp7	1314,6	1744,8	252,3	**2199,1**
		wp8	2255,7	1808,6	172,7	**2896,4**
		wp9	1928,0	832,7	189,5	**2108,7**
		wp10	2033,6	764,4	215,5	**2183,2**
		wp11	1691,6	1504,9	269,9	**2280,1**
		mean:	1555,7	960,1	203,9	**1877,4**
		stdev:	798,6	632,4	88,7	**943,4**

LEMD2LEBL (CV.k=10): input=FP(3D)+AP → output=Lat/Lon/Alt (NN)

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)

DART SESAR JOINT UNDERTAKING

- Stage-2: HMMs vs. Linear Regressors:
  - ➢ LR improves accuracy (9-34%) vs. HMM across all configurations
  - ➢ clustering/LR better (up to 16,5%) than clustering/HMM
- Advanced non-linear regressors, generalization (vs. LR, CART):
  - ➢ More resilient to "noise" in training, better generalization & stability
  - ➢ Clustering (stage-1) becomes less important with more robust regressors
  - ➢ NN: seems the best tradeoff between complexity vs. performance

Compared to other state-of-the-art approaches (up to 2019):

❖ S. Ayhan, H. Samet, "Aircraft trajectory prediction made easy with predictive analytics", Proc. ACM SIGKDD (2016)
  - ➢ **RMSE (3-D): 12 km** (HMM-based, using enrichments, no use of flight plans)

❖ Y. Liu, M. Hansen, "Predicting aircraft trajectories: a deep generative convolutional recurrent neural networks approach" Tech. rep., ITS,Univ. Calif. (2018), arXiv:1812.11670
  - ➢ **RMSE (3-D): 92 km** (CNN-based, using enrichments + flight plans)

DATA STORIES

Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS RESEARCH CENTER

Data Science Lab @ Univ. of Piraeus

URL: http://datastories.org
Facebook: @DataStories
Twitter: @UnipiDataSciLab

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Track & Know

datAcron

DART

Ref: H. Georgiou, N. Pelekis, S. Sideridis, D. Scarlatti, Y. Theodoridis (2019) “Semantic-aware Aircraft Trajectory Prediction using Flight Plans”, International Journal of Data Science Analytics (IJDSA), Mar.2019 (doi:10.1007/s41060-019-00182-4).