A Spatio-Temporal Track Association Algorithm Based on Marine Vessel Automatic Identification System Data

Welcome to the online companion of our paper. There are three subfolders located under the root folder along with this README file. The contents of these three subfolders are explained below:

Dataset:

Since the challenge datasets used in this study are not owned by us, we do not include them in this folder. They can be accessed from the Gitlab repository of the data challenge competition [1]. Our R scripts are supposed to load them directly from the Gitlab repository to generate association results and save them in the "Result" subfolder. If this repository [1] is removed or changed in the future, please contact us for the datasets.

This subfolder stores only the reduced version of the additional AIS dataset that we used for the performance comparison. The full dataset can be downloaded from the AIS data repository [2]. To reduce the dataset, we only selected the first 25000 rows of the dataset. Also, to make it similar to the challenge datasets, we remove the "heading", "ship type" and "destination" column from the original dataset.

Codes:

This subfolder contains all the scripts to generate results in the paper. The primary scripting language used is R. In R, we used several packages apart from the basic packages and they are required to be installed and loaded before running any of the R code (.r) files. The required packages are listed below:

- Geosphere
- Rcurl
- lubridate
- ggmap
- ggplot2

However, to generate the multi-object trackers (MOT) result files (for comparison), we used Matlab and the related codes are contained in another subfolder titled "MOT Codes" inside the Codes folder. "Matlab Sensor Fusion and Tracking Toolbox" needs to be installed to run these files on Matlab. Under the Code folder, we have several R scripts. Their purposes are explained below:

OnlineClustering.r: This is the primary R script implementing our online clustering algorithm for the 5 test datasets under the ATD challenge [1]. It will load the test datasets contained in the Gitlab repository [1], run the association algorithm and generate the track association .csv file with a separate "clusterID" column. This result will be then passed to the posthoc algorithm for merging clusters generated in this step.
Posthoc.r: This R script will generate different dataframes which will help us to merge online clusters generated in the previous step according to our posthoc algorithm described in the paper. The final cluster files will be saved in the "Result" folder for performance analysis.
OnlineClusteringAdditionalAIS.r: This R script will generate the online clustering result for the reduced version of the additional AIS dataset [2] ("Dataset" folder) used for comparison purposes.

PosthocAdditionalAIS.r: This R script will help merge the clusters generated in the previous step for the additional AIS dataset.

Table II.r: This R script will reproduce Table II of the paper. It will fetch the final cluster files after posthoc merging saved in the "Result" subfolder. For each test instance, it will generate the missed, extra, merged, broken and swapped tracks following the approach suggested by the data challenge organizers. It will generate the performance measure for the sample algorithm as well. The final result files for the sample algorithm can also be found in the "Result" subfolder.

Table III.r: This R script will reproduce Table III in the paper. It will fetch the final cluster files after posthoc merging saved in the "Result" subfolder. For each test instance, it will generate the continuity scores, mean and median completeness scores following the approach suggested by the data challenge organizers. It will generate the performance measure for the sample algorithm as well. The final result files for the sample algorithm can also be found in the "Result" subfolder.

Table IV.r: This R script will reproduce Table IV in the paper. It will fetch the final cluster files after posthoc merging saved in the "Result" subfolder. For each test instance, it will generate the number of clusters generated by our algorithm and sample algorithm.

Table V-VI.r: This R script will reproduce the performance metric of MOT reported in Table V and Table VI of the paper. It will fetch the final cluster files of MOT saved in the "Result" subfolder. For each test instance, it will generate missed, extra, merged, broken and swapped tracks following the approach suggested by the data challenge organizers.

Table VII.r: This R script will reproduce the performance metric of MOT reported in Table VII in the paper. It will fetch the final cluster files of MOT saved in the "Result" subfolder. For each test instance, it will generate the continuity scores, mean and median completeness scores following the approach suggested by the data challenge organizers.

Table VIII.r: This R script will reproduce the performance metric of our algorithm and the sample algorithm for an additional AIS dataset reported in Table VIII of the paper. It will fetch the final cluster files saved in the "Result" subfolder. Then, it will generate all the performance measures following the approach suggested by the data challenge organizers.

Figure 7.r: This R script will reproduce Figure 7 in the paper.

Result:

This subfolder contains all the final cluster files generated using our algorithm, sample algorithm and MOT using the test datasets stored in the Gitlab repository [1] and additional AIS dataset.

Machine Specification: Intel Core i7(7700HQ@2.80 GHz), 16GB Ram; Windows 10

Software Version: R version 3.6.2; MATLAB version 2020b

Reproducing the results in the paper:

For the convenience of the user of this online companion, in the following table, we summarize how to reproduce different tables and figures used in the paper.

Which Results to Reproduce	Data File	Code File	Output
Table II	Final cluster files saved in the "Results" subfolder. Which are: "finalresulttestset1.csv", "testset2finalresult.csv", "finalresulttest3.csv", "finaltest4.csv", "reallyfinaltestset5.csv", "AIS_Challenge_Problem_Set_1(8tracks)_notional.csv", "AIS_Challenge_Problem_Set_2_notional.csv", "AIS_Challenge_Problem_Set_3_notional.csv", "AIS_Challenge_Problem_Set_4_notional.csv", "AIS_Challenge_Problem_Set_5_notional.csv"	Table II.r	All the contents of Table II
Table III	Same as above	Table III.r	All the contents of Table III
Table IV	Same as above	Table IV.r	Last 2 rows of Table IV
Table V	Final cluster files saved in the "Results" subfolder using MOT codes. Which are: "result22.csv", "result23.csv", "result24.csv", "result25.csv", "result26.csv", "result27.csv", "result28.csv", "result29.csv", "result30.csv", "result31.csv"	Table V-VI.r	3rd and 4th row of Table V
Table VI	Same as above	Table V-VI.r	3rd and 4th row of Table VI
Table VII	Same as above	Table VII.r	3rd and 4th row of Table VII
Table VIII	Final cluster files saved in the "Results" subfolder using the additional AIS dataset. Which are: "resultreview325correctedmain.csv" and "review3_train2_notional.csv"	Table VIII.r	All the contents of Table VIII
Figure 7	Final cluster files used to generate Table II	Figure 7.r	Figure 7

[1] D. Mercer. Algorithms for Threat Detection. Accessed: Oct. 28, 2020. [Online]. Available: https://gitlab.com/algorithms-for-threat-detection/2019/atd2019

[2] Marine Traffic Automatic Identification System (AIS): Open Research Maritime Vessel Tracking Dataset. Accessed: Jan. 16, 2022. [Online]. Available: https://www.marinetraffic.com/research/dataset/marinetraffic-automatic-identification-system-ais/

**Thank you for using this online companion. If you have any questions on implementing our algorithm, please feel free to send an email at imtiaz.ahmed@mail.wvu.edu or imtiaz_avi@yahoo.com