Imitation Learning for ship collision avoidance using AIS Data

Submitted by: **Saad Momin**

Roll no: **NA20B041**

Guided by: **Dr. Abhilash Sharma**

Department of Ocean Engineering

B. Tech Project Report

Signature of the student

Signature of the advisor

Timeline

October 23	Literature Review and Topic Finalisation
October 23	Primary AIS Data Analysis of AIS Data
November 23	Data Visualisation and Pre-processing
November 23	Built the framework for Encounter Identification
December 23	Write-up and Submission of Report
January 24	Using Polars Library for parameter Tuning

1 Introduction

This project is divided into two main parts: first, to identify conflicting trajectories through massive AIS data and second, to leverage that information to imitate the RL model, addressing the collision avoidance problem. This report mainly focuses on the first part of the project, which is the extraction of conflicting trajectories .

In recent years, the growing volume of maritime traffic has posed a challenge to navigational safety. Researchers have sought to improve maritime safety by analyzing historical collision data; however, this approach is hindered by the limited number of incidents collected within a given period. The Autonomous Identification System (AIS) has made an enormous amount of maritime traffic data available freely. AIS tracks vessel movement through onboard transceivers and terrestrial and satellite-based stations. The data collected by AIS contain broadcast kinematic and static information, facilitating ship movement tracking and providing opportunities for discovering maritime traffic knowledge, including movement behavior analysis, route estimation, and detecting anomalous behaviors. This research aims to identify potential ship traffic conflicts using AIS data.

Traffic conflict refers to trajectories that could lead to a collision if ships do not take evasive action. In other words, conflicting trajectories can be treated as near-collision cases for analysis. In order to achieve our goal, we required a robust and efficient method. Drawing inspiration from framework proposed by Po-Ruey Lei on his study on *"Mining maritime traffic conflict trajectories from a massive AIS data"*, we developed a framework aimed at discovering conflicting trajectories associated with each encounter a ship undergoes.

The framework comprises three steps:
1. **Window-Based Encounter identification**: Grouping ship trajectories that might intersect.
2. **Conflict Ascertainment**: Detecting possible conflicts between from identified encounters.
3. **Conflict Trajectory(CT) Generation**: Collecting all identified conflicts chronologically to form trajectories.

This approach will serve as a foundation for our future work, specifically in training a reinforcement learning (RL) model. The CTs obtained will play a crucial role in enhancing the RL model's effectiveness for ship collision avoidance by acting as an expert.

The report is organized as follows: **Section 2** contains information about AIS data and preliminary observations made on experimental data. **Section 3** outlines the technical details of the proposed framework. **Section 4** describes the results of the proposed scheme when applied to a real data set collected from an AIS system. Conclusions and future works are discussed in **Section 5**.

2 Properties of AIS Database

This section describes the attributes of AIS data and the preliminary analysis done on AIS data, which will help us understand the data and get the most information out of it.

2.1 AIS Data Attributes

The automatic identification system (AIS) is widely used to prevent maritime traffic collisions. AIS technology broadcasts ship information and voyage details at regular intervals, typically every 2 to 10 seconds if the ship is underway or every 3 minutes if the ship is at anchor. This information can be received by onboard transceivers and terrestrial and/or satellite base stations continuously over a distance of 20 nautical miles (approx 37 kilometers).

Several crucial attributes are associated with AIS data, including the vessel's maritime mobile service identity (MMSI), base date time, longitude, latitude, speed over ground (SOG), course over ground (COG), vessel type, vessel dimensions, rate of turn (ROT), navigation status, and heading. In this project, we have only considered the more relevant attributes for conflicting trajectory estimation. They are ship unique identity (MMSI), base date time, ship geographical position (longitude and latitude), SOG, and COG.

2.2 Preliminary Data Analysis

We started with organizing the data for ease of manipulation. We sorted it first in increasing time order and subsequently by the unique MMSI for each ship. This arrangement ensures that the trajectory of each individual ship is presented in chronological order, facilitating streamlined processing.

In Figure 1, we showcase an extracted set of AIS trajectory data covering a maritime area of 16 by 12 km offcost of Florida in North Atlantic Ocean. The dataset encompasses a notable number of trajectories and an impressive 15,373,253 spatial-temporal points.

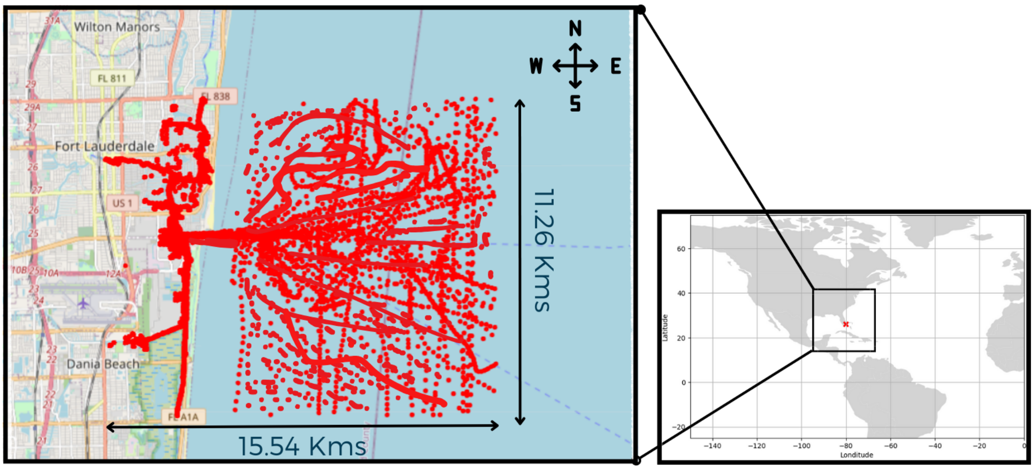

Figure 1: Spatial distribution of AIS Data covering area of 16-by-12 Kms approx

Figure 2, shows the temporal distribution of datapoints over the period of 6 months having unifrom distribution over all.

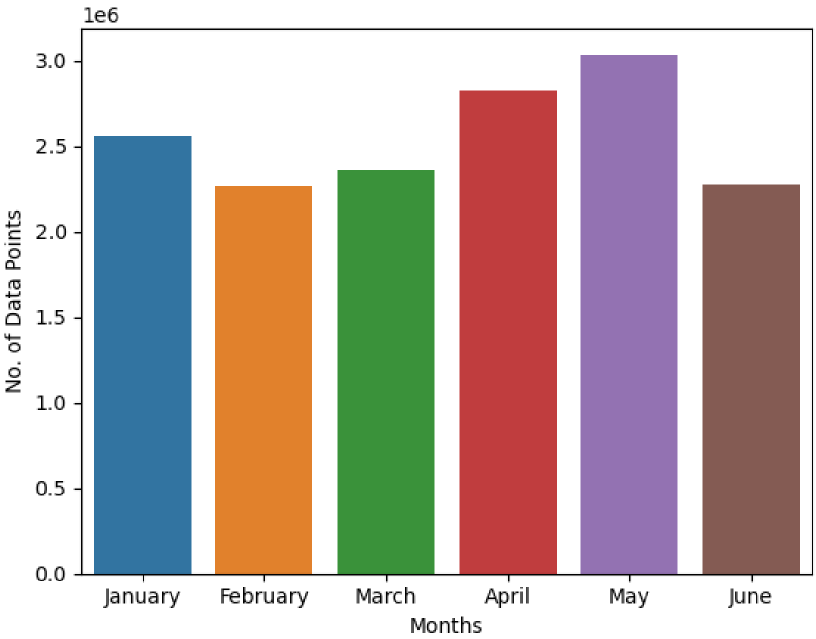

Figure 2: Temporal distribution of AIS Data over a period of six months

Where as Figure 3, indicates distribution of Speed Over Ground which is speed of the vessel relative to the surface of the Earth. It is typically measured in knots and Course Over Ground which is direction in which the vessel is moving relative to the surface of the Earth. It is typically measured in degrees from true north, clockwise. Since our data is collected from near portside region more than 90% of ships have slow speed same can be seen from pi chart.

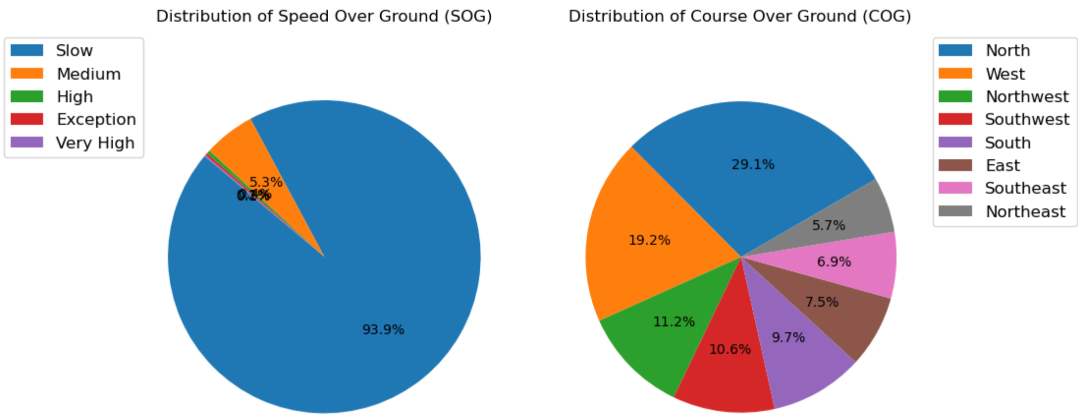

Figure 3: Distribution of Speed Over Ground (SOG) and Course Over Ground (COG)

Figure 4 shows the categorical classification of continuous values of COG and SOG.

Course over ground (COG)	Statuses
[337.5, 360]∪ [0, 22.5)	North
[22.5, 67.5)	Northeast
[67,5, 112.5)	East
[112.5, 157.5)	Southeast
[157.5, 202.5)	South
[202.5, 247.5)	Southwest
[247.5, 292.5)	West
[292.5, 337.5)	Northwest

Speed over ground (SOG)	Statuses
[0, 3)	Slow
[3, 14)	Medium
[14, 23)	High
[23, 99)	Very high
Over 99	Exception

Figure 4: List of COG and SOG status

2.3 Data pre-processing and data cleaning

Over the 6-month period, a ship may undergo different navigation routs/paths, making it important to identify specific trajectories for training our Imitation Learning model in future. We came up with a method to categorize different trajectories: ”A vessel is considered to be under navigation conditions if the Speed Over Ground (SOG) value is not equal to zero,” meaning that continuous sequences of AIS data with nonzero SOG values are marked as trajectories. Furthermore, to ensure the usefulness of a trajectory, we disregarded those with short sequences lacking sufficient data points for training and testing. Our threshold was set at 100 AIS data points. After applying these conditions, we identified 2,459 unique trajectories, constituting 404,344 data points.

The temporal data we have is not uniformly spaced in time due to variability in AIS data transmission intervals, which depend on whether a ship is underway (with intervals of 2 to 10 seconds depending on speed) or anchored (with intervals of 3 minutes). For ease of data handling, it is important for all data points to have an equal time interval. Therefore, we opted for a 1-minute time interval and interpolated or extrapolated the data.

The final uniformly time-spaced data consisted of **2,342,553** data points.
The data showed abnormal SOG values (such as 120 knots) and numerous instances of SOG jumps across the data, which required close attention, as can be seen in Figure 5. We calculated the ship’s distance traveled over a given time period and compared it to the actual distance between those two points to see if the SOG jumps were valid or if there was an error. If there was a noticeably large deviation, we changed the SOG value to the last value. The same procedure was followed for SOG with exceptional values, such as 120 knots.

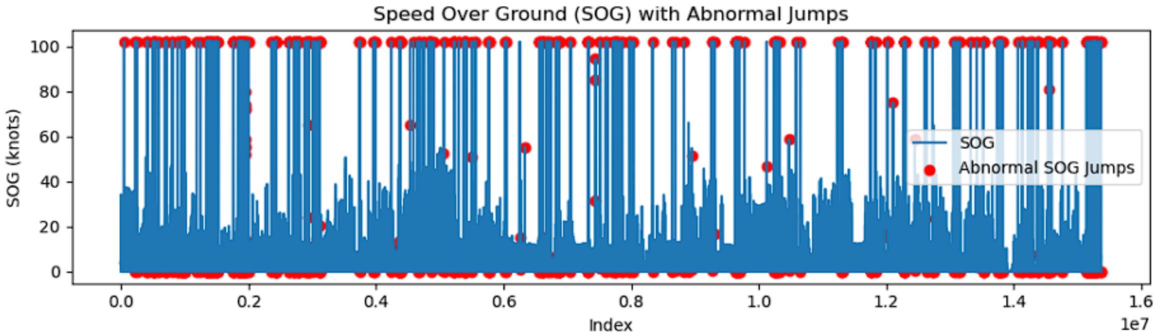

Figure 5: SOG values with abnormal Jumps

3 Discovery of Cluster of Conflicting Trajectories (CCT)

This section outlines the three-step approach for determining the cluster of conflicting trajectories.

1. Estimating the Encounter
2. Determine the Conflicting Encounter
3. Aggregating the Conflicting Encounter to form Conflicting trajectories (CT)

In order to find the CT, we mined the trajectories for conflicting scenarios. As a ship navigates its trajectory, it engages with numerous other vessels. When an encounter transpires with another ship, the involved ship takes corresponding actions. If such encounters have the potential to result in a collision if no action were taken, we categorize them as conflicting situations.

To determine the encounters of a parent ship with its neighboring vessels, we assume each ship to have a circle of observation with a user-defined radius, denoted as d_0. Subsequently, we assess whether other ships fall within this observation circle. If no ships are detected, we conclude that the ship is not currently in an encountering condition. However, if other ships fall within the observation circle, we deem the ship to be in an encountering condition.

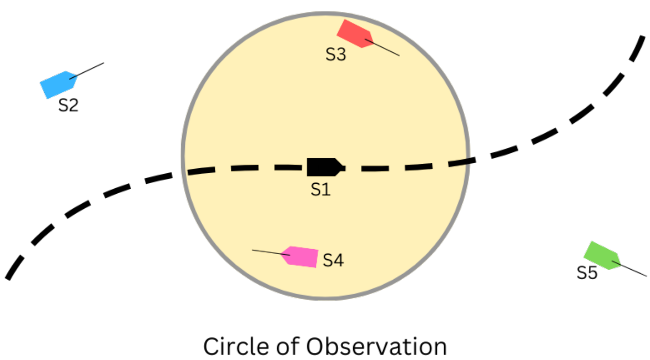

Figure 6: Ship under Encounter condition with circle of observation

Having identified encounters throughout the trajectories, our goal is to pinpoint those that lead to conflicts. To achieve this, we leverage the concepts of Time to Closest Point of Approach (TCPA) and Distance to Closest Point of Approach (DCPA), which will be elaborated upon in an upcoming section. By aggregating trajectories during conflicting time durations, we can derive the conflicting trajectories that a ship may encounter during its voyage.

3.1 Estimating the Encounter

To detect the conflicting trajectories, the first task is to discover the groups of ships associated with encounter situations. A ship is deemed to be in an encountering condition if it is sailing in close proximity to another ship or a group of ships.

We begin by considering a trajectory of a ship and its circle of observation with radius d_0 as mentioned earlier. At a given observation time t_i and a collected AIS trajectory data D_A, we can extract a snapshot of data $D(t_i) = \{s_{i_1}^{t_i}, s_{i_2}^{t_i}, \ldots, s_{i_n}^{t_i}\}$, where $s_{i_n}^{t_i} = (x_{i_n}^{t_i}, y_{i_n}^{t_i}, v_{i_n}^{t_i}, c_{i_n}^{t_i})$ represents the geoposition of s_{i_n}, and $v_{i_n}^{t_i}$ and $c_{i_n}^{t_i}$ indicate the speed and course, respectively. Subsequently, we determine if any ship falls within the circle of observation of the parent ship. If so, we mark this ship as encountering at time t_i, else we go to next timestamp.

Given the large AIS dataset, looking over all the data at each time stamp could pose computational challenges. To address this, we propose an efficient and novel window-based approach. In this approach, the data space is initially divided using a geometric square grid structure. The grid size, denoted as Cw. The window is then constructed using a 3-by-3 cell-based configuration. As illustrated in **Figure 7**, only data within the cell-based window is considered in the computation of closeness for encounter clustering. This means that the trajectory points of ships separated by long distances are disregarded. This approach significantly reduces computational costs, ensuring an effective and streamlined process.

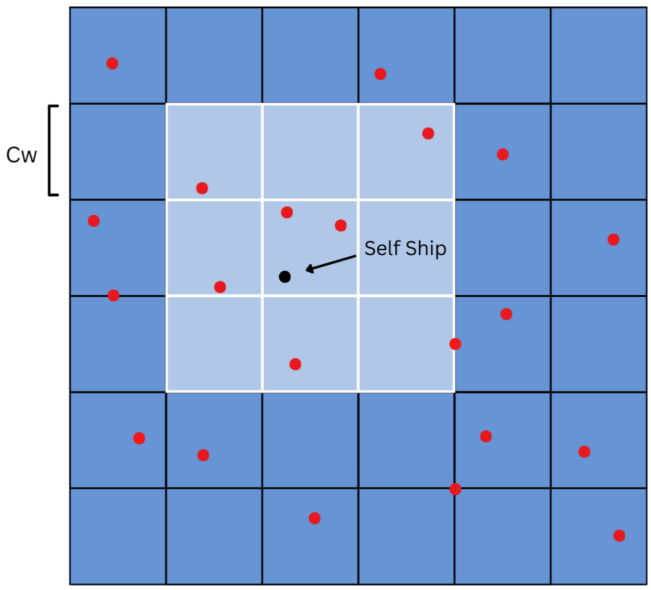

Figure 7: Data grid at time ti with window

3.2 Determining Conflict

After the window-based encounter estimation, the next step involves discovering conflicting encounters. As mentioned earlier, maritime conflict behavior is characterized by two crucial aspects: Distance to Closest Point of Approach (DCPA) and Time to Closest Point of Approach (TCA). In the conflict ascertainment phase, DCPA and TCA are utilized to assess the maritime conflict behaviors within encounter situations.

Given a pair of ships s_1 and s_2 at observation time t_i, we can derive their mobility information as $s_1^{t_i} = (x_1^{t_i}, y_1^{t_i}, v_1^{t_i}, c_1^{t_i})$ and $s_2^{t_i} = (x_2^{t_i}, y_2^{t_i}, v_2^{t_i}, c_2^{t_i})$. The variation in their positions over time can be formulated as $(x_1(t), y_1(t)) = (x_1 + v_1.x.t, y_1 + v_1.y.t)$ and $(x_2(t), y_2(t))$ = $(x_2 + v_2.x.t, y_2 + v_2.y.t)$. Kinematics can be used to estimate the distance between s_1 and s_2 as follows:

$$\text{Dist}_{s_1,s_2}(t) = \sqrt{\text{Dist}_{(s_1,s_2)}(t).\mathbf{x} + \text{Dist}_{(s_1,s_2)}(t).\mathbf{y}} \tag{1}$$

where,

$$\begin{aligned}\text{Dist}_{(s_1,s_2)}(t).\mathbf{x} &= [(x_1 + v_{1.x}.t) - (x_2 + v_{2.x}.t)]^2 \\ \text{Dist}_{(s_1,s_2)}(t).\mathbf{y} &= [(y_1 + v_{1.y}.t) - (y_2 + v_{2.y}.t)]^2\end{aligned}$$

Then, by employing kinematics, DCPA and TCA can be derived as:

$$DCPA(s_1, s_2) = \min \text{Dist}_{(s_1,s_2)}(t) \tag{2}$$

$$\begin{aligned}TCA(s_1, s_2) &= \arg\min \text{Dist}_{(s_1,s_2)}(t) \\ &= -\frac{(x_1 - x_2)(v_{1.x} - v_{2.x}) + (y_1 - y_2)(v_{1.y} - v_{2.y})}{(v_{1.x} - v_{2.x})^2 + (v_{1.y} - v_{2.y})^2}\end{aligned} \tag{3}$$

Depending on the value of TCPA, we can have three different types of scenarios:

1. If $TCPA > 0$: The DCPA will take place in future.
2. If $TCPA = 0$: The occurrence of DCPA is happening.
3. If $TCPA < 0$: DCPA has already occurred.

The encounter is designated as conflicting if any two members $s_1^{t_i}$ and $s_2^{t_i}$ satisfy both conditions $DCPA \leq d_c$ and $TCA \geq 0$, where d_c is a user-defined conflicting distance (taken as 1 km).

3.3 Clustering Conflicting Trajectories

In the final phase of the proposed framework, we aggregate conflicting encounters at each observation time to derive trajectories depicting conflicting scenarios (refer to Figure 8 for an illustrative example). Within a navigation sequence, we pinpoint three timestamps C_{t_i}, C_{t_i+1}, C_{t_i+2} of conflicting situations. These instances are accumulated to form a trajectory $CT = \{C_{t_i}, C_{t_i+1}, C_{t_i+2}\}$, denoted as a conflicting trajectory.

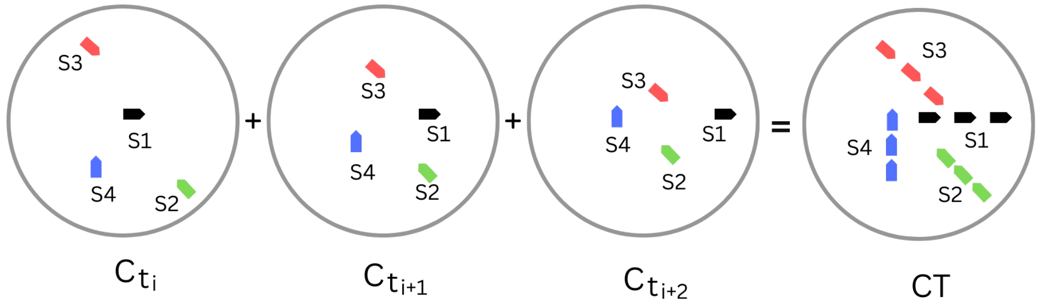

Figure 8: Example of conflicting trajectory generation

The number of timestamps may vary based on distinct scenarios. Similarly, with K conflicting situations timestamps, $CT = \{C_{t_i}, C_{t_i+1}, C_{t_i+2}, ...C_{t_i+k}\}$. These trajectories may be considered as the set of ship movement behavior in near-collision situations. Through a comprehensive analysis of potential collisions and interactions among diverse trajectories extracted from maritime traffic data, our ultimate objective is to leverage this knowledge to train a reinforcement learning (RL) model capable of real-time collision avoidance.

4 Experiments and Results

In accordance with the proposed framework, the window-based encounter estimation identified a total of 405,174 encounters from the modified AIS trajectories containing 2,342,553 data points . Subsequently, we aimed to detect conflicts associated with each discovered encounter. Only 49 conflicting encounters detected. Finally, 10 conflicting trajectories were constructed by aggregating these identified conflicting encounters in chronological order of observation time. All these conflicting trajectories occur in the 18th hour of the day.

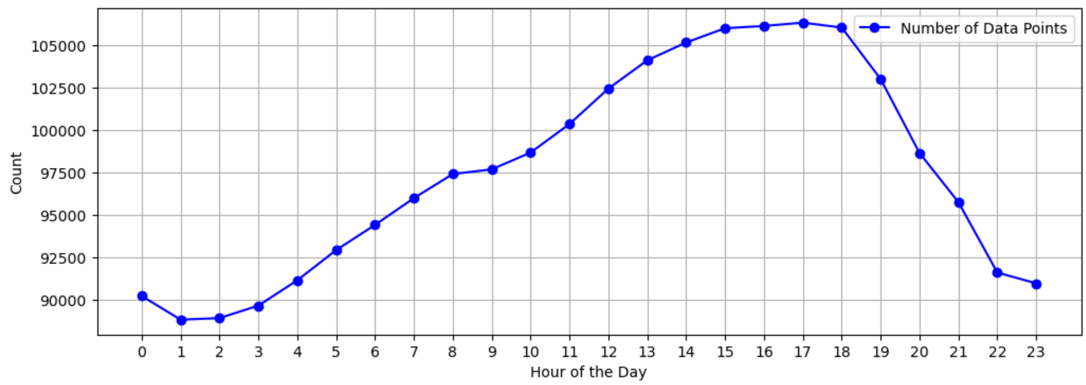

Figure 9: Number of Data Points at Each Hour of the Day

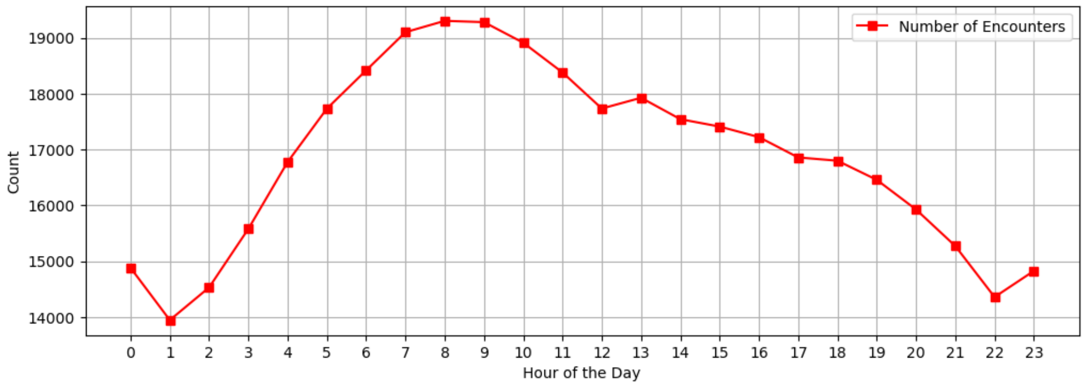

Figure 10: Number of Encounters at Each Hour of the Day

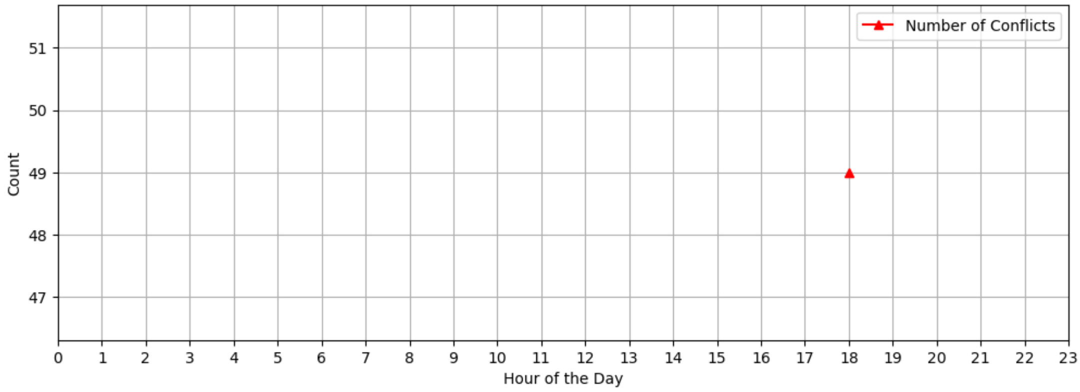

Figure 11: Number of Conflicts at Each Hour of the Day

Figure 12 demonstrates a case of discovered conflicting trajectories. Fig. 12a presents the initial cluster of ships S1, S2, and S3 with a possible conflict; the arrows indicate the locations of ships moving toward each other. In Fig. 12b, the conflict condition of S1 and S2 is illustrated, with their movements in near-collision locations marked by a red circle. From Fig. 12c, we can observe S2 staying in position while S1 moves ahead, avoiding the collision. Fig. 12d shows another conflict encountered by S1 with S3, marked with a red circle, and in Fig. 12e, the historical trajectories of all ships are presented.The Animation of the demonstrated case can be seen with this Link : https://drive.google.com/file/d/1CEXa3SM2knKIUxx3lKTSFMLTz55OLr2I/view?usp=sharing.

The results of the case demonstration indicate that the proposed framework is able to effectively discover conflicting trajectories from collected maritime AIS data.

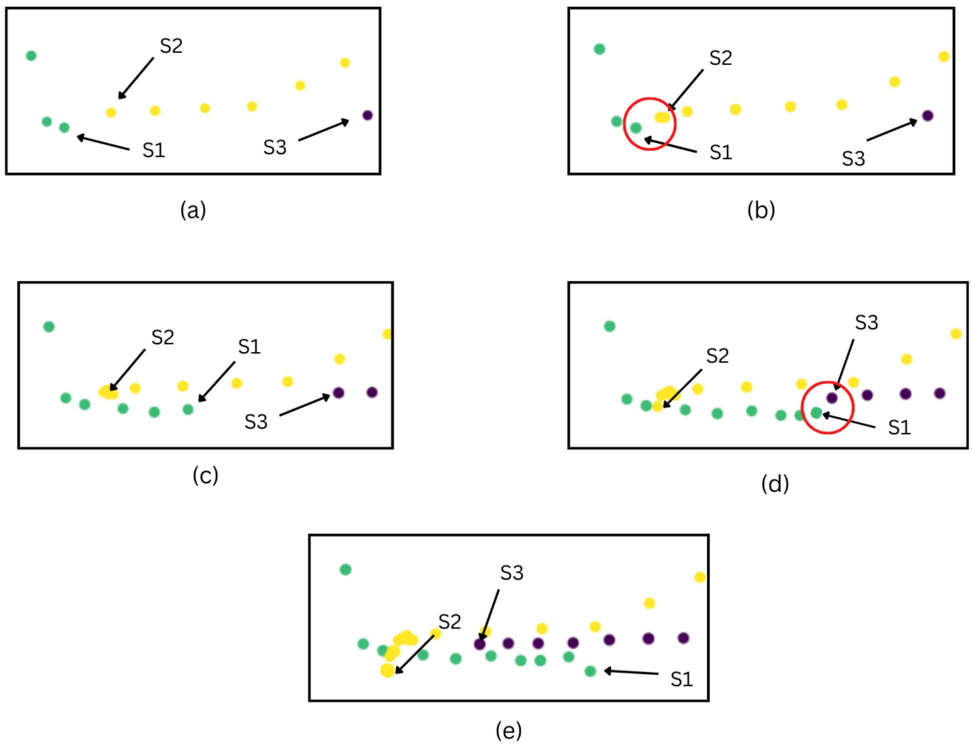

Figure 12: Case Demonstration of Discovered CT

5 Conclusion and Future Work

The study concludes with the successful development of a framework utilizing AIS data for the discovery of conflicting ship trajectories, crucial for collision prevention. Implemented in three phases—encounter estimation, conflict detection, and CT generation—the framework demonstrates adaptability to the complexity of AIS data.

For future work, hyperparameter tuning can be explored for variables such as the radius of the circle of observation (d_0), the size of the grid cell ($n \times n$), the length of the grid cell (C_w), and their impact on the number of encounters, conflicts, and execution time.

Additionally, the study aims to leverage identified conflicting trajectories to train a reinforcement learning (RL) model using imitation learning (specifically behavioral cloning). This model will consider various evasive actions in response to conflicts, thereby enhancing collision avoidance strategies.

References

[1] Po-Ruey Lei : Mining maritime traffic conflict trajectories from a massive AIS data

[2] Shangbo Mao1, Enmei Tu, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally, and Guang-Bin Huang: An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining

[3] Enmei Tu, Member, IEEE, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally,and Guang-Bin Huang : Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology