Local Anomaly Detection In Maritime Traffic Using Visual Analytics

Fernando H. O. Abreu
Dalhousie University
Halifax, NS, Canada
fernando.abreu@dal.ca

Amilcar Soares
Memorial University of Newfoundland
St. Johns, NL, Canada
amilcarsj@mun.ca

Fernando V. Paulovich
Dalhousie University
Halifax, NS, Canada
paulovich@dal.ca

Stan Matwin
Dalhousie University
Halifax, NS, Canada
stan@cs.dal.ca

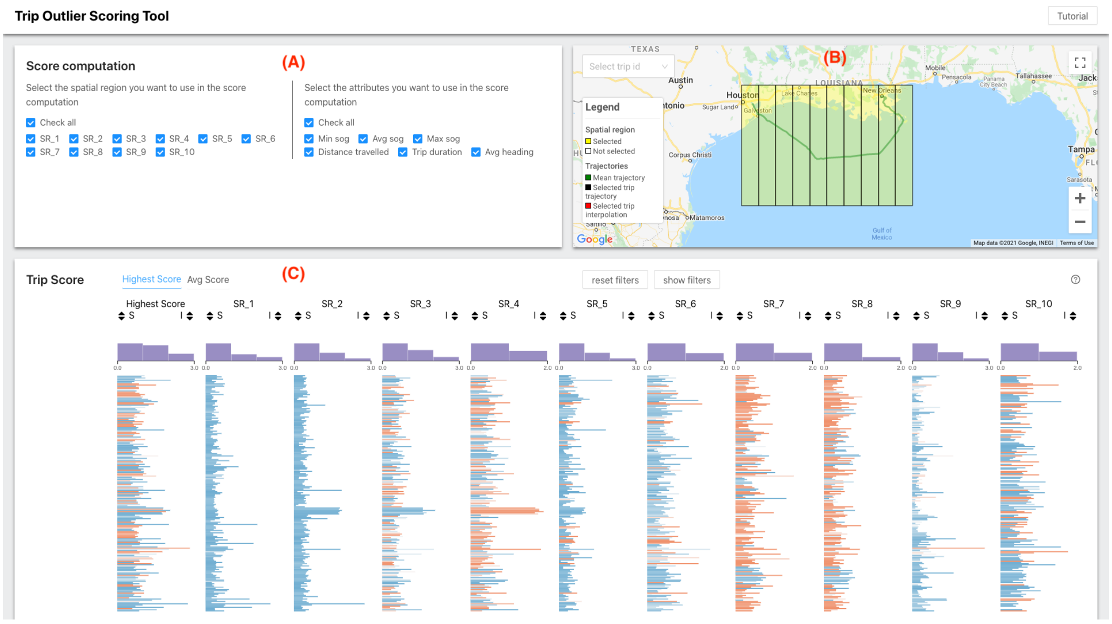

Figure 1: Overview of the Trip Outlier Scoring Tool (TOST). The user uses the Score computation component (A) to control which spatial regions and attributes will be used in the score. The trip scores are visualizes in the Trip Score component (C) where the user can filter and sort the data, and select a trajectory trip to be displayed in the map (B).

ABSTRACT

With the recent increase in sea transportation usage, maritime surveillance's importance to detect unusual vessel behavior related to several illegal activities has also risen. Unfortunately, the data collected by the surveillance systems are often incomplete, creating a need for the data gaps to be filled using techniques such as interpolation methods. However, such approaches do not decrease the uncertainty of ship activities. Depending on the frequency of the data generated, they may even confuse operators, inducing them to errors when evaluating ship activities to tag them as unusual. Using domain knowledge to classify activities as anomalous is essential in the maritime navigation environment since there is a well-known lack of labeled data in this domain. In an area where finding which trips are anomalous is a challenging task when using solely automatic approaches, we use visual analytics to bridge this gap. In this work, we propose a tool that uses spatial regions to divide trips into *subtrajectories* and score them. The scores are displayed in a tabular visualization where users can rank trips by segment to find local anomalies. The amount of interpolation in *subtrajectories* is displayed together with scores, and the trip is displayed on the map so users can use their insight to make sense if the score is reliable.

1 INTRODUCTION

Maritime transportation is essential nowadays; about 90 percent of everything traded in the world is done by sea [11].Since 2004, vessels of 300 gross tonnages or more which travel internationally, and cargo ships of 500 gross tonnages or more are obligated by the International Maritime Organization (IMO) to have Automatic Identification System (AIS) onboard[1] which produces a constant high volume of data [14]. This technology transmits the vessel destination, speed, position, and many other items of static information, such as ship name and Maritime Mobile Service Identity (MMSI), which is used to identify a ship uniquely [11].

[1] http://www.imo.org/en/OurWork/Safety/Navigation/Pages/AIS.aspx

The Department of Defense of Canada (DRDC) and surveillance authorities, such as Coastal Marine Security Operation Centres (MSOCs) which are responsible for guaranteeing coastal safety, have an interest in using this data to uncover several potential issues [5], such as illegal transport of drugs, human trafficking, fishing in illegal areas, illegal immigration, sea pollution, piracy, and even terrorism [1]. These activities have a significant impact on society, environment, and economy, and for such, it is essential to identify these types of events as soon as possible [16]. Vessels involved in these types of illegal activities usually follow specific patterns like unexpected stops, speeding, and deviations from standard routes [1, 11]. Ships that are operating legally commonly travel through the same route due to regulations and because it is usually the shortest path between ports, which would decrease the vessel fuel consumption. For this reason, ships that navigate non-standard routes or show signals of route deviations can be potentially labeled as presenting anomalous behavior [1]. However, identifying which trips are anomalous is not an easy task for maritime operators due to the large volume of data produced by AIS systems, which creates an overload of instances to be analyzed manually. Currently, operators usually use systems that display vessels on a world map that they can use to track their movements [6]. Although this can help operators reach some awareness of what is going on in the sea, it can prove a difficult task trying to identify anomalous vessels among a large number of normal vessels [5].

Many works focus on finding anomalies in an automated manner, such as [7], [11] and [20] which use different clustering techniques to extract a group of trajectories with similar behavior. Then other methods are used to classify the trajectories. However, the problem of automatically identifying anomalies is very complex and not well-defined [13]; additionally, it requires dynamic adaptation since humans will always try to change their *modus operandi* to not get caught, which in turn, makes automatic systems less reliable [12]. Thus, systems that automatically detect anomalies are rarely used in the real world [12, 13]. On the other hand, visualizations make use of humans' inherent ability to perceive patterns and filter information in combination with their creativity and background knowledge [8, 13], which allows them to be able to analyze and understand complex, massive, and dynamic data.

Some known works in the field, such as [13] and [5] use a combination of visualization and automated techniques to aid the user when trying to identify anomalies. However, the vast majority of algorithms proposed to identify anomalies automatically may not work for *local anomalies* [18], or they require labeled data to train a model [4, 15]. This means that deviations from normality that happen just in a small portion of a vessel trajectory may be left out when considering the trajectory as a whole, especially when analyzing works in the maritime domain. The only work we found that could partially address this issue is [17]. Their method chooses N equally spatially distributed sample points for trips, and then it classifies them as anomalous routes with low probabilistic density points. However, this work may miss local anomalies depending on the number of samples chosen, while ours use all trajectory points. Their tool only works for positional data, while we use several attributes.

Lastly, when analyzing vessel trajectories from raw AIS data, it can be faulty and incomplete, and it can happen for multiple reasons. First, one of the frequencies used by AIS transceivers is Very High Frequency (VHF), which makes AIS data unreliable [19]. Second, Vessel Traffic Service (VTS) stations may miss several AIS messages from vessels traveling close to the coast due to information overloading [10]. Third, even though Satellite AIS has become more common since it can capture longer ranges than shore-based AIS, it is common for the data received by it to have gaps. Finally, there are also cases where vessel crew interfere with AIS signal or turn the transponder off to cover illegal activities [9]. For this reason, vessel trajectories often need to be interpolated, which can increase algorithm accuracy [3]. However, the interpolated data's anomalies may be incorrect if the interpolation was not done correctly or when many consecutive data points are missing. Therefore, it would be important to present information related to interpolation if an anomaly is detected in the interpolated region of a trajectory, such as what was the quality of that interpolation or show the interpolation itself, so one can assess if the interpolation was done properly and if it is indeed an anomaly. The user could also further investigate what could have happened when there was no signal. However, to our knowledge, there is no work in this field that allows users to explore the potential impact of interpolation on anomalies.

In this paper, we propose a tool that aims to tackle the problems mentioned above. We make very few assumptions about who the users of this tool could be. This paper contributes with the proposal and development of a visual analytics tool for finding local anomalies in trip trajectories while also taking into account the trip's interpolation. Section 2 describes the proposed tool and discusses some of the decisions that were made. Section 3 we show a use case of our tool. Finally, in Section 4, we present a summary of this work and discuss some of our tool's limitations; and we propose some ideas for future work.

2 TRIP OUTLIER SCORING TOOL (TOST)

As mentioned previously, this work aims to develop a tool for identifying local anomalies in trip trajectories while also providing users some information about the interpolation, such as where and how it happened and how much interpolation there is on the trajectory. In this work, a trip is defined by the sequence of a vessel's AIS messages when traveling from one port to another. A *spatial region* can be defined as a 2-dimensional geographic polygon. In this work, we create it automatically for the user by creating a minimal box containing all points of all trajectories that traveled between two specific ports and then divide it into N spatial regions of same area. Finally, a *subtrajectory* is a sequence of points of a trajectory contained in a *spatial region.*

Figure 2 shows an overview of our framwework's steps. It is composed of a preprocessing step that combines two sources of AIS data to get *trips'* information. Trips that don't share the same origin and destination are removed. The remaining trips go through a cleaning process where invalid data, such as outlier points, are removed, and gaps are interpolated. We then create spatial regions that serve the purpose of partitioning each trip *trajectory* into *subtrajectories.* The *subtrajectories*' attributes, such as average speed, is given a score based on how much they deviate from the mean over all other trips attribute values; the combined final score for each *subtrajectory* is then displayed in a tabular visualization. Each trip is represented as a row in the table where the first column may show the maximum or average score for a trip, depending on the user's selection. The other columns show the *subtrajectory* scores, which are represented by a bar length, while the color of the bar shows the amount of interpolation in the *subtrajectory.*

We first display an overview of the overall maritime situation in the table. The users can then use filters to remove uninteresting data, so it shows only trips of interest. They can hover or select an individual row to see the scores and interpolation values of a trip. By clicking on a row, the trajectory trip will be displayed on the map. The user can then compare the trajectory trip against the mean trajectory to see if there were any deviations and if the interpolation was done correctly. The user can also choose which attributes and spatial regions should be used during the score computation, which will update the *subtrajectory* score.

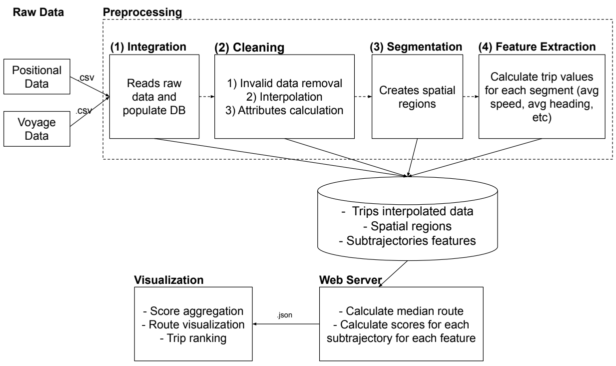

Figure 2: Overview of the framework of the Trip Outlier Scoring Tool

Our tool has three main components: the Score computation (A), a map (B), and Trip Score table (C), as shown in Figure 1. The Score computation allows the users to chose which spatial regions and attributes they want to use to compute the scores for each trip *subtrajectory*. As an aggregate final score for each *trip*, we may show the highest score, which is the highest value amongst all trip *subtrajectories*, or it can show the average score of the trip *subtrajectories*. In order to calculate a *substrajectory* score, we first calculate the z-score for each attribute selected by the user. Then these values are summed together and divided by the number of attributes. When calculating a subtrajectory attribute z-score, the population comprises all other *subtrajectories* created by the same spatial region for trips with the same origin and destination ports.

The Map was created to display the previously created regions as well as trip trajectories. It is displayed with a zoom on the region containing the two ports. Since we want the user to differentiate the original *points* and from the ones that were created after the interpolation, we distinguish them by color. The black portion of the trajectory was created from the original data points, while the red portion was interpolated {colorblueas can be seen in Figure 5. We also display a mean trajectory in the map, representing a path that a trip should make. This trajectory is calculated using a function of the tool created by Erland et al. [2].

In the Score Table each row in this table represents a trip. For each column, there is a bar in which its length represents the *subtrajectory* aggregated score, and the color represents the percentage of interpolated points. The bar's height is dynamic; they change based on how many trips are being displayed at a given time. A longer bar may indicate a higher deviation from normality since our score is derived from the z-score. Longer bars also stand out in comparison to smaller bars. And the interpolation is displayed as a gradient from blue to red. The exact scores and interpolation values for a trip, as well as the trip id, can be seen at the bottom of a table when a user hovers over a row with the mouse. At the top of the table, we show the distribution of each region's scores as purple bars. This visualization has two purposes: first, the user can brush the region to filter out uninteresting vessels, and so decreasing the number of vessels displayed at the table which could improve the table visibility. Second, showing the distribution may reveal a spatial region with a higher number of outliers than others or a region where the outliers have a much higher score.

3 A USE CASE

In this use case, we exemplify the use of TOST[2] for finding speed anomalies far from shore. The dataset used includes trips of cargo ships that traveled from Houston to New Orleans from 2009 to 2018. We first use the Score Computation (see Figure 1(A)) to select only regions 5, 6, and 7, and we selected only the average speed attribute that is the main target of this analysis. Other options for regions could have also been used by clicking on the yellow regions on the map (see Figure 1(B)). If the user clicks on those controls, these interactions would recompute the scores and update the visualization only to display the regions of interest.

Next, we choose to have the first column to display by highest score or average score. Since we want to highlight trips that may have an outlier behavior, we chose the one with the highest score even in only a single region. Given that many trips are being displayed, we filter out trips with a score below 2.5 by brushing the score distribution in the Highest Score column. This could also have been accomplished by inputting this value manually after clicking "show filters", which is useful when high precision is necessary, the updated trip score table can be seen in Figure 3. By looking at the filtered trips, we can see that most subtrajectories have some degree of interpolation, especially in region 7, which may indicate that it is a region where the terrestrial tower cannot capture the AIS messages.

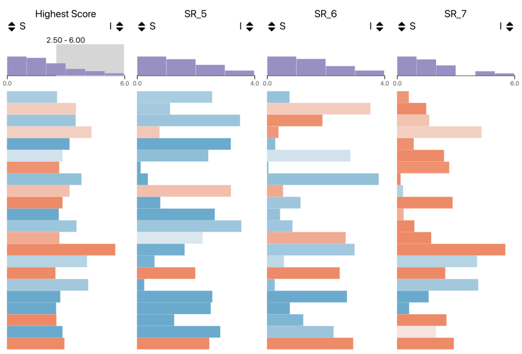

Figure 3: Trip scores filtered to show only trips with score above 2.5

After, we rank the trajectories by the highest score and hover the mouse on top of the row to see the trip's scores, which has the subtrajectory with the highest score. This score belongs to the trip with id equals to 2187, as can be seen in Figure 4. Trip 2187 has a high score, especially on region 6 and 7. We can also see that in region 7, all points are interpolated, which indicates that this score is not reliable since the region is not has a considerable size. If we click on the row to plot this trip trajectory in the map, we can see that this interpolation does not seem reliable; thus,

[2]https://gitlab.com/Fernando-Abreu/thesis_project

the score for this subtrajectory cannot be trusted. After plotting, the expert should think if this gap size makes sense or if this trip needs further investigation.

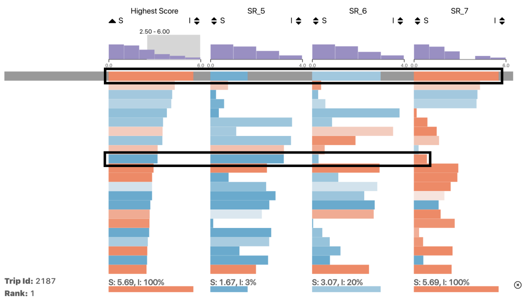

Figure 4: Trip Scores with trip with highest subtrajectory score selected. Trips ranked 1 and 10 are highlighted

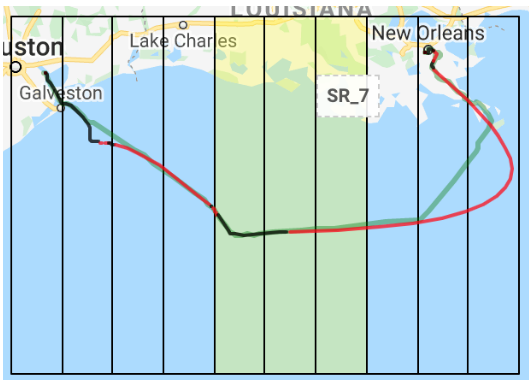

Figure 5: Trip 2187 trajectory

Another example is trip 339, which is on rank 10 of our selection. When we look at the table, we can see that although the tool added some interpolated points on subtrajectories in regions 6 and 7, region 5 had an outlier behaviour. When we hover this row to see that it had a 0 percent interpolation and score of 3.28. Therefore, this score is very reliable, and the user could frame this as an outlier behavior. If the expert decides to have a close look at the data, they could see that this trip had an average speed of 5.93 knots in region five, while the average speed in that particular region is 15.69 knots with a 3.24 standard deviation. Now it is the expert's job to try to understand why the vessel navigated so slowly in that region compared to other vessels. The conclusion of the investigation could point to engine issues or unregulated or illegal activity associated with the vessel.

4 CONCLUSION

In this work, we identified local anomalies using a combination of features and used an interpolation strategy to give the user a certain degree of reliability to the anomaly. We achieved this goal by proposing and developing a web tool that partitions and scores each *subtrajectory* regarding its attributes. Users can interact with this tool through filtering and sorting to find trips with local anomalies. They can also plot trajectories trips in the map and identify which portions of that trajectory were interpolated.

Future works include using a clustering algorithm to group trips with similar trajectories to compare the same class of vessels and have a more fine-grained analysis. We also intend to add a page that allows the users to choose between creating the spatial regions automatically or manually. If the user chooses to create manually, the user should be able to draw spatial regions on a map using drawing tools in the map. Otherwise, the tool will create regions based on trajectory patterns or using trajectory segmentation methods.

REFERENCES

[1] Enrica d'Afflisio, Paolo Braca, Leonardo M Millefiori, and Peter Willett. 2018. Detecting anomalous deviations from standard maritime routes using the Ornstein–Uhlenbeck process. *IEEE Transactions on Signal Processing* 66, 24 (2018), 6474–6487.

[2] Willem Eerland, Simon Box, Hans Fangohr, and András Sóbester. 2017. Teetool–a probabilistic trajectory analysis tool. *Journal of Open Research Software* 5, 1 (2017).

[3] Dini Oktarina Dwi Handayani, Wahju Sediono, and Asadullah Shah. 2013. Anomaly detection in vessel tracking using support vector machines (SVMs). In *2013 International Conference on Advanced Computer Science Applications and Technologies*. IEEE, 213–217.

[4] Amílcar Soares Júnior, Chiara Renso, and Stan Matwin. 2017. Analytic: An active learning system for trajectory classification. *IEEE computer graphics and applications* 37, 5 (2017), 28–39.

[5] Valérie Lavigne. 2014. Interactive visualization applications for maritime anomaly detection and analysis. In *ACM SIGKDD Workshop on Interactive Data Exploration and Analytics*. 75.

[6] Etienne Martineau and Jean Roy. 2011. *Maritime anomaly detection: Domain introduction and review of selected literature.* Technical Report. DEFENCE RESEARCH AND DEVELOPMENT CANADA VALCARTIER (QUEBEC).

[7] Steven Mascaro, Ann E Nicholso, and Kevin B Korb. 2014. Anomaly detection in vessel tracks using Bayesian networks. *International Journal of Approximate Reasoning* 55, 1 (2014), 84–98.

[8] Lucas May Petry, Amilcar Soares, Vania Bogorny, Bruno Brandoli, and Stan Matwin. 2020. Challenges in Vessel Behavior and Anomaly Detection: From Classical Machine Learning to Deep Learning. In *Advances in Artificial Intelligence*, Cyril Goutte and Xiaodan Zhu (Eds.). Springer International Publishing, Cham, 401–407.

[9] Fabio Mazzarella, Michele Vespe, Alfredo Alessandrini, Dario Tarchi, Giuseppe Aulicino, and Antonio Vollero. 2017. A novel anomaly detection approach to identify intentional AIS on-off switching. *Expert Systems with Applications* 78 (2017), 110–123.

[10] Van-Suong Nguyen, Nam-kyun Im, and Sang-min Lee. 2015. The interpolation method for the missing AIS data of ship. *Journal of Navigation and Port Research* 39, 5 (2015), 377–384.

[11] Giuliana Pallotta, Michele Vespe, and Karna Bryan. 2013. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy* 15, 6 (2013), 2218–2245.

[12] Maria Riveiro and Göran Falkman. 2011. The role of visualization and interaction in maritime anomaly detection. In *Visualization and Data Analysis 2011*, Vol. 7868. International Society for Optics and Photonics, 78680M.

[13] Maria Riveiro, Göran Falkman, Tom Ziemke, and Håkan Warston. 2009. VISAD: an interactive and visual analytical tool for the detection of behavioral anomalies in maritime traffic data. In *Visual Analytics for Homeland Defense and Security*, Vol. 7346. International Society for Optics and Photonics, 734607.

[14] Amílcar Soares, Renata Dividino, Fernando Abreu, Matthew Brousseau, Anthony W Isenor, Sean Webb, and Stan Matwin. 2019. CRISIS: integrating AIS and ocean data streams using semantic web standards for event detection. In *2019 International Conference on Military Communications and Information Systems (ICMCIS)*. IEEE, 1–7.

[15] Amílcar Soares, Jordan Rose, Mohammad Etemad, Chiara Renso, and Stan Matwin. 2019. VISTA: A visual analytics platform for semantic annotation of trajectories.. In *EDBT*. 570–573.

[16] Iraklis Varlamis, Ioannis Kontopoulos, Konstantinos Tserpes, Mohammad Etemad, Amilcar Soares, and Stan Matwin. 2020. Building navigation networks from multi-vessel trajectory data. *GeoInformatica* (2020). https://doi.org/10.1007/s10707-020-00421-y

[17] Guizhen Wang, Abish Malik, Calvin Yau, Chittayong Surakitbanharn, and David S Ebert. 2017. TraSeer: A visual analytics tool for vessel movements in the coastal areas. In *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*. IEEE, 1–6.

[18] Wanqi Yang, Yang Gao, and Longbing Cao. 2013. TRASMIL: A local anomaly detection framework based on trajectory segmentation and multi-instance learning. *Computer Vision and Image Understanding* 117, 10 (2013), 1273–1286.

[19] Daiyong Zhang, Jia Li, Qing Wu, Xinglong Liu, Xiumin Chu, and Wei He. 2017. Enhance the AIS data availability by screening and interpolation. In *2017 4th International Conference on Transportation Information and Safety (ICTIS)*. IEEE, 981–986.

[20] Rong Zhen, Yongxing Jin, Qinyou Hu, Zheping Shao, and Nikitas Nikitakos. 2017. Maritime anomaly detection within coastal waters based on vessel trajectory clustering and Naïve Bayes Classifier. *The Journal of Navigation* 70, 3 (2017), 648.