A real-time trajectory classification module

Ioannis Kontopoulos[*]
Department of Informatics and Telematics, Harokopio University of Athens
Greece
kontopoulos@hua.gr

Antonios Makris
Department of Informatics and Telematics, Harokopio University of Athens
Greece
amakris@hua.gr

Konstantinos Tserpes
Department of Informatics and Telematics, Harokopio University of Athens
Greece
tserpes@hua.gr

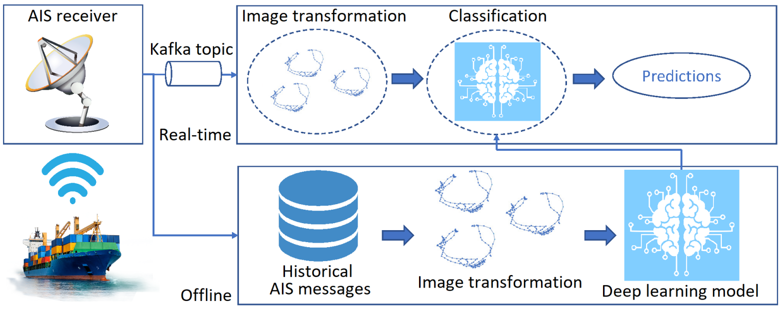

Figure 1: The architecture of the trajectory classification module.

ABSTRACT

Nowadays, massive volumes of mobility data are being generated from thousands of tracking devices, such as GPS devices, RFID sensors, location-based services, satellites, and wireless communication technologies. This phenomenon can be strongly observed in the maritime domain and as a result, today's industry is flooded with tracking data originating from vessels across the globe that transmit their position at frequent intervals. Automated methodologies able to extract meaningful information and identify mobility patterns from such tracking data are of utmost importance since they can reveal abnormal or illegal vessel activities in due time. To this end, we present a demo of a trajectory classification methodology that is able to classify vessels' trajectories into activities that the vessels are engaged in from AIS data streams in real-time. The goal is to provide maritime authorities with a visualization tool and an API of the vessel trajectories and their activities in real-time. The trajectory classification methodology that is used in this demo achieves a classification performance of over 95%.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; **Computer vision**; • **Human-centered computing** → *Visualization application domains.*

[*]Corresponding author.

EMODE '23, November 13, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0347-8/23/11.
https://doi.org/10.1145/3615885.3628005

KEYWORDS

trajectory classification, spatio-temporal analysis, AIS, deep learning

ACM Reference Format:

Ioannis Kontopoulos, Antonios Makris, and Konstantinos Tserpes. 2023. A real-time trajectory classification module. In *1st ACM SIGSPATIAL International Workshop on Methods for Enriched Mobility Data: Emerging issues and Ethical perspectives 2023 (EMODE '23), November 13, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3615885.3628005

1 INTRODUCTION

The maritime industry generates large volumes of data through tracking sensors, particularly the Automatic Identification System (AIS) used for vessel location. Originally designed for safety, AIS is now also utilized by maritime authorities to identify illegal activities and monitor vessel behavior, enhancing Maritime Situational Awareness (MSA). This increased awareness is essential for various purposes, including detecting unusual vessel behavior due to damage, aiding in Search And Rescue (SAR) operations, collision prevention, combating piracy, and addressing illegal fishing activities.

Researchers have developed techniques that use vessel trajectory data to enhance Maritime Situational Awareness (MSA). These algorithms focus on features like trajectory clustering, classification, anomaly detection, and event prediction. Trajectory classification is a key method for identifying abnormal patterns or events. Such techniques allow the authorities to further take advantage of the trajectories formed from the AIS messages either in real-time [3, 6] or on historical data [11, 12]. Several approaches for trajectory classification attempt to exploit global features such as mean velocity, acceleration, or their standard deviations [1, 7]. Other approaches try to apply trajectory partitioning [5] or find relevant sub-trajectories [2] in an attempt to identify more discriminative features. In most of these studies, the context of the analysis is typically the physical world and the geography. The three fundamental features of a possibly multidimensional space are: position, speed and acceleration. Experts, however, heavily rely on trajectory visualization to identify and categorize trajectories that are of some significance. This provides an intuition to move the analysis in a different domain, by leveraging computer vision approaches (i.e. Convolutional Neural Networks (CNNs)) on trajectory classification. One of the most common goals of such networks is to classify a set of images into a predefined set of labels of interest.

In this work, we present a demo of a real-time, trajectory classification methodology for vessels that employs a deep learning streaming methodology over AIS data streams. The methodology for this demo was presented in [3, 4]. Streaming vessel trajectory patterns are transformed into images, which in turn are accurately classified into vessel activities in real-time, using deep neural networks. Our system relies on the well-known Lambda architecture [10] in an attempt to balance latency and throughput for the 4Vs of the Big Data (i.e., Volume, Velocity, Veracity, and Variety). A 'batch-processing' layer (offline) is responsible for training the classification model, which is then used to distinguish the vessel activities in the 'stream-processing' layer (real-time) at which data continuously arrive from vessels at high rates.

2 ARCHITECTURE

Several frameworks exist for distributed stream processing, namely Apache Spark[1], Apache Flink[2] and Kafka streams[3]. Out of these three options, Apache Spark is not preferred since it performs micro-batching over streams of events and a system is needed that can handle real-time event processing. Despite the fact that both Apache Flink and Kafka streams have inherent support for real-time stream processing, the development of the trajectory classification module was limited to the frameworks' programming paradigm. Therefore, to balance event processing with low latency, high throughput, and the ability to freely develop the module that would be optimized for our use case, the Apache Kafka[4] framework was used. Apache Kafka is a distributed publish-subscribe and message-exchange platform similar to a message queue able to process streams of events as they arrive. The Apache Kafka ecosystem consists of topics, producers, and consumers. A kafka topic is a category/feed name to which messages are stored and published. A producer is an application that continuously publishes or stores messages on a topic. A consumer is an application that is subscribed to a topic and continuously consumes messages. A kafka topic can be divided into k partitions with each partition storing different messages. Specifically, messages with the same key will be stored in the same partition. k consumers can be subscribed to the partitioned topic with each consumer reading from a different partition thus enabling high throughput. A producer can store messages to the partitioned topic and Apache Kafka will handle the load balancing of the messages among the partitions internally. In our use case, the vessel identifier can be considered as the message key, the AIS receiver as the producer, and the trajectory classification module as the consumer. An even distribution of load within the nodes of the system reduces the probability that a node turns to a hotspot and its property also acts as a safeguard to the system reliability [8, 9].

[1]https://spark.apache.org/streaming/
[2]https://flink.apache.org/
[3]https://kafka.apache.org/documentation/streams/
[4]https://kafka.apache.org/

The architecture of the trajectory classification module can be broken down into three parts as seen in Figure 1. The first part (left of Figure 1) refers to the collection of streams of AIS messages via a base station installed in the premises of the Department of Informatics and Telematics, at Harokopio University of Athens[5]. The base station is comprised of a Very High Frequency (VHF) antenna that is connected to a Raspberry Pi. Then, an NMEA-0183[6] multiplexer is used, called Kplex[7], to receive messages from the antenna. Then, these messages are decoded with the use of the DMA AisLib Java library[8], are sent through a Kafka topic and stored in a PostgreSQL[9] database.

The second part (bottom of Figure 1) refers to the training of a machine learning model. Specifically, AIS messages stored in the database are annotated to specific labels or activities, namely anchored, moored, fishing, and underway. Then, the trajectories of the vessels engaged in such activities are transformed into images. To do so, and because the distance each vessel travels through space is different (e.g. vessels traveling in the Atlantic ocean will cover greater distances compared to vessels traveling in the Irish sea), the bounding box or the surveillance area in which the vessel moves needs to be normalized. In order to accurately capture and place the shape of the trajectory inside a normalized bounding box, the total distance of both the x and the y axis in which the vessel moves must be defined first. Therefore, we calculate the total horizontal distance and the total vertical distance the vessel travels based on the minimum and maximum longitudes and latitudes, respectively. Then, the distance each AIS position has traveled from the minimum longitude and latitude is calculated. Next, the percentage of the total distance each AIS position has traveled from the minimum coordinate in both x and y axes is also calculated. As a result, each AIS position can be placed inside a normalized bounding box or a surveillance space of a user-defined size that is essentially an image representation. More details can be found in [3, 4]. Finally, a well-established, deep learning model, called VGG16 is trained on the annotated trajectories that have been transformed into images. VGG16 is pre-trained on the ImageNet[10] dataset and its weights are updated based on the new training image dataset created from the AIS messages.

The third part (top of Figure 1) is the real-time classification of the streaming AIS trajectories, where messages are collected through the AIS receiver and the Kafka topic and transformed into images at run-time. Then, the already trained VGG16 model is used to classify the newly created images and predict an annotation for the vessel. To classify the vessels' trajectories, the module uses a temporal sliding window of a 6-hour length and a step of one AIS message. Therefore, in every received message the module

[5]https://dit.hua.gr/index.php/en/
[6]An electrical and data specification for communication between marine electronics such as GPS receivers. It has been defined and is controlled by the National Marine Electronics Association (NMEA).
[7]https://github.com/stripydog/kplex
[8]https://github.com/dma-ais/AisLib
[9]https://www.postgresql.org/
[10]https://www.image-net.org/

takes into account all of the AIS messages in the past 6 hours and converts them into an image. The six-hour window was chosen because it is long enough for patterns to emerge and short enough to classify these patterns as soon as possible. Furthermore, previous experiments [3] demonstrated that the six-hour window is the optimal window for better classification results. Next, the deep learning model classifies the image and outputs a probability for each of the predefined vessel activities. The vessel activity with the highest probability is the final prediction of the module. It should be noted that the classification is performed on a per-vessel basis and that multiple classification models can run at the same time and in a distributed fashion to support larger volumes of data streams and number of vessels due to Apache Kafka. Currently, the classification module consumes approximately 266, 112 messages per day from one AIS base station and from over 1200 vessels in the area of the Saronic Gulf of Attica, Greece. The classification module can be seen live at HUA AIS Station. It is also worth noting that the training of the deep learning model is trained in Keras[11], Python and the real-time processing of AIS messages and classification is implemented in Java's Deeplearning4j[12] library.

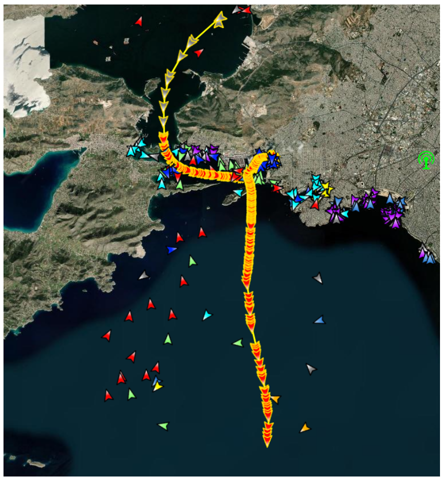

Figure 2: A vessel's past 6-hour trajectory.

3 MODULE FUNCTIONALITIES

At first glance, the user is able to visualize a vessel's past 6-hour trajectory by clicking on a vessel of choice. Figure 2 demonstrates this use case. Different colors of vessel markers indicate a different vessel type, e.g., a passenger vessel. A gray vessel marker indicates that the vessel has not yet transmitted static messages that contain information about its vessel type. Static messages are transmitted once every six minutes and include the vessel's static information such as type, draught, and flag. Dynamic messages are transmitted from once every three seconds and up to once every three minutes depending on the vessel's speed and rate of change of course. In the example of Figure 2, it can be observed that the vessel's marker is initially colored gray and then becomes red which indicates a tanker vessel.

Moreover, the user can see information regarding its activities when clicking on the selected vessel. Figure 3 illustrates an example of a vessel that has been classified as engaged in fishing. When fishing vessels are engaged in fishing activities, they tend to manoeuver. Due to the fact that other types of vessels could manoeuver (e.g., anomalous behavior) the label is mentioned as manoeuvering and not fishing in order to be more generic. The user can decide based on the type of the vessel, the exact activity that it is engaged in. For visualization purposes, a pop-up appears that illustrates the probability for each pre-defined activity or label. The activity with the highest probability is also visualized as the final classification of the vessel's activity for that point in time and space. Also, the pop-up illustrates other information regarding the vessel such as its name, identification number, and speed as seen in Figure 3.

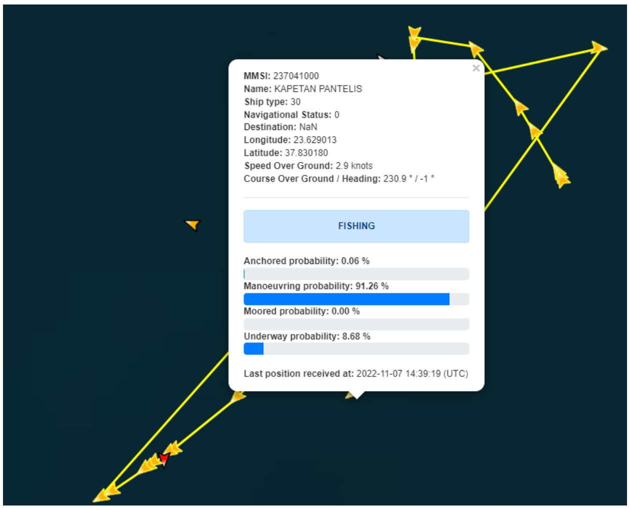

Figure 3: Example of a vessel classified as being engaged in fishing.

Furthermore, the trajectory classification module is able to visualize the coverage of the AIS base station. When the button "Receiver Coverage" is clicked, a blue polygon appears on top of the map that indicates the geographic boundaries of the receiver's coverage. In essence, the polygon is the convex hull of all the AIS messages received in the past four hours. Given a set of points in space, the convex hull is the smallest convex polygon that contains all the points.

Finally, the user can also access the data illustrated in the classification module via a REST API which will appear when the "Data Access" button is clicked. Through this API, the user can retrieve information that includes a vessel's past twelve-hour trajectory, a vessel's trajectory during a specific period in time, and all vessels' trajectories received from the AIS station during a specific period in time.

[11]https://keras.io/
[12]https://deeplearning4j.konduit.ai/

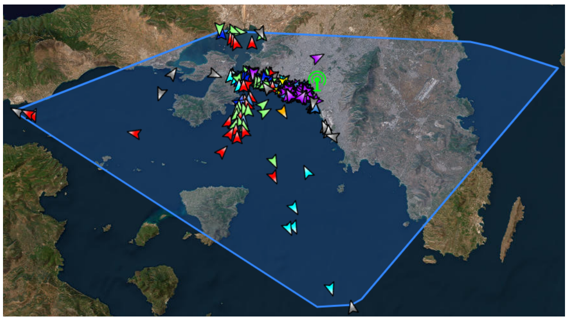

Figure 4: Network coverage.

Table 1: Results of the evaluation of the trajectory classification module.

Methodology	Cross Validation		
	Precision	**Recall**	**F1-score**
First Dataset	0.9755	0.9705	0.9723
Second dataset	0.9686	0.9652	0.9653

4 EXPERIMENTAL EVALUATION

This paper presents a demo and a visualization tool of a trajectory classification methodology. This section briefly presents the classification performance of the methodology used in the demo and it has already been presented in more detail in previous works [3, 4]. Two datasets were used for experimental evaluation of the trajectory classification module. The first dataset contained AIS messages collected from a Terrestrial AIS receiver (T-AIS) that covers the Saronic Gulf (Greece) and contains high-quality AIS information without gaps of information. The vessels have been monitored for almost one and a half month period starting on February 18$^{\text{th}}$, 2020, and ending on March 31$^{\text{th}}$, 2020. The dataset provides information for 1229 unique vessels and contains $11,769,237$ AIS records in total. The AIS messages used for our ground truth dataset contain activities that have been extracted from vessels engaged in the following activities: underway, anchored and moored. The second dataset that was used was provided by MarineTraffic[13] and contains AIS messages from January 1st, 2018 to February 28th, 2018 in the seas of Northern Europe. The AIS messages used for our ground truth dataset contain the following activities: trawling, longlining, moored, and underway. The total number of AIS messages in this dataset sums up to $61,050$. In both datasets, the navigational status, the vessel type, and the destination reported in the AIS messages were used as annotations for the labels. Furthermore, manual annotation and filtering also took place to create representative samples for training.

To evaluate the trajectory classification module, we performed a 10-fold cross-validation on both datasets, keeping at each fold 90% of the data for training and 10% of the data for validation and reported the macro-average results. The evaluation results demonstrate the high-accuracy classification performance of the trajectory module and are illustrated on Table 1.

[13]https://www.marinetraffic.com

5 CONCLUSION

In this work, we presented a demo of a methodology that is able to classify vessels' trajectories in real-time using a deep neural network. This demo module can provide maritime experts with insight regarding vessels' activities and supports several functionalities that highlight the past and current behavior of multiple vessels in a specific region. Data consumed by the AIS receiver are publicly available via a REST API. Future works include adding support for more vessel activities such as tugging. Finally, due to the unparalleled quantities of trajectory data, which in turn can overwhelm human analysis approaches, we intend to utilize compression techniques in order to minimize the size of the trajectory data, while at the same time minimizing the impact on the trajectory analysis methods.

ACKNOWLEDGMENTS

This work was supported by the SmartShip and MASTER Projects through the European Union's Horizon 2020 research and innovation program under Marie-Sklodowska Curie Grant Agreement Nos. 823916 and 777695, respectively. The work reflects only the authors' view, and the EU Agency is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] Camila Leite da Silva, Lucas May Petry, and Vania Bogorny. 2019. A survey and comparison of trajectory classification methods. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, 788–793.

[2] Carlos Andres Ferrero, Luis Otavio Alvares, Willian Zalewski, and Vania Bogorny. 2018. Movelets: Exploring relevant subtrajectories for robust trajectory classification. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. 849–856.

[3] Ioannis Kontopoulos, Antonios Makris, and Konstantinos Tserpes. 2021. A deep learning streaming methodology for trajectory classification. *ISPRS International Journal of Geo-Information* 10, 4 (2021), 250.

[4] Ioannis Kontopoulos, Antonios Makris, Dimitris Zissis, and Konstantinos Tserpes. 2021. A computer vision approach for trajectory classification. In *22nd IEEE International Conference on Mobile Data Management, MDM 2021, Toronto, ON, Canada, June 15-18, 2021*. IEEE, 163–168. https://doi.org/10.1109/MDM52706.2021.00034

[5] Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez. 2008. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. *Proceedings of the VLDB Endowment* 1, 1 (2008), 1081–1094.

[6] Chun-Xun Lin, Tsung-Wei Huang, Guannan Guo, and Martin DF Wong. 2018. MtDetector: a high-performance marine traffic detector at stream scale. In *Proceedings of the 12th ACM International Conference on Distributed and Event-Based Systems*. 205–208.

[7] Antonios Makris, Camila Leite da Silva, Vania Bogorny, Luis Otavio Alvares, Jose Antonio Macedo, and Konstantinos Tserpes. 2021. Evaluating the effect of compressing algorithms for trajectory similarity and classification problems. *GeoInformatica* 25, 4 (2021), 679–711.

[8] Antonios Makris, Konstantinos Tserpes, and Dimosthenis Anagnostopoulos. 2017. A novel object placement protocol for minimizing the average response time of get operations in distributed key-value stores. In *2017 IEEE International Conference on Big Data, BigData 2017, Boston, MA, USA, December 11-14, 2017*. IEEE Computer Society, 3196–3205. https://doi.org/10.1109/BigData.2017.8258300

[9] Antonios Makris, Konstantinos Tserpes, Dimosthenis Anagnostopoulos, and Jörn Altmann. 2017. Load balancing for minimizing the average response time of get operations in distributed key-value stores. In *14th IEEE International Conference on Networking, Sensing and Control, ICNSC 2017, Calabria, Italy, May 16-18, 2017*. IEEE, 263–269. https://doi.org/10.1109/ICNSC.2017.8000102

[10] Nathan Marz and James Warren. 2015. Big Data: Principles and best practices of scalable realtime data systems. *Manning Publications* (2015).

[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] Dimitris Zissis, Konstantinos Chatzikokolakis, Giannis Spiliopoulos, and Marios Vodas. 2020. A distributed spatial method for modeling maritime routes. *IEEE Access* 8 (2020), 47556–47568.