Computer Science and Information Systems 21(4):1963–1978 https://doi.org/10.2298/CSIS240301063G

Maritime Trajectory Mining: An Automatic Zones of Interests Discovery and Annotation Framework*

Omar Ghannou, Etienne Thuillier, and Omar Boucelma

Aix-Marseille University, CNRS, LIS
Marseille, France
{omar.ghannou, etienne.thuillier, omar.boucelma}@univ-amu.fr

Abstract. As global traffic continues to grow, the identification of areas of particular significance, known as Zones of Interest (ZOI), becomes crucial for optimizing transportation systems and analyzing mobility patterns. In the maritime domain, effective ZOIs discovery is essential for enhancing route planning, improving safety measures, and managing resources efficiently. Within the context of trajectory mining, these ZOIs provide valuable insights into movement behaviors and operational efficiencies. In this paper, we present a framework for discovering and annotating ZOIs within maritime trajectories. The proposed approach involves processing raw positional data to initially identify candidate ZOIs, which are subsequently refined using contextual information. By leveraging real georeferenced vessels trajectories, collected from thousands of commercial ships, this framework proposes a structure of elements that will be implemented as part of the TNTM French project. While this research contributes to maritime field by providing a method for ZOIs discovery and annotation, it can be generalized to various application domains that may leverage of mobility data analytics.

Keywords: Maritime Trajectory Mining, Zone Of Interest, ZOI, Area Of Interest, AOI, Stops Extraction, Classification, Contextual Approach, OpenStreetMap, OSM, VGI

1. Introduction

Trajectory mining or trajectory data mining is an interdisciplinary research area that focuses on analyzing and extracting knowledge from trajectory data, which comprises spatio-temporal information collected from moving objects, such as GPS logs, mobile phone signals, and surveillance cameras. Trajectory mining has been widely applied in various domains, such as transportation, urban planning, social network analysis, or environmental monitoring.

For the specific maritime transportation domain, as global traffic continues to grow, accounting for over 80% of transportation transactions worldwide [6], the characterization of vessel trajectories is an activity of major importance.

In this paper, we focus on specific points of a trajectory, that is, Zones of Interest (ZOI). A ZOI denotes a geographical area that attracts particular attention for various reasons, such as significant events, research endeavors, or specific actions aligned with stakeholder priorities.

* This paper is an extended paper of the published DC paper "Automatic Discovery of Zones of Interests with Maritime Trajectory Mining" in ADBIS2023 conference

In the maritime domain, ZOIs are essential for optimizing routes, assessing risks, and identifying commercial opportunities. For instance, maritime carriers can leverage ZOI knowledge to evaluate risks related to weather conditions or illicit activities and capitalize on commercial prospects like servicing booming ports or supplying raw materials.

The increasing availability of GPS data and the proliferation of Location-Based Services have enabled the extraction of valuable insights from trajectory data. Effective ZOI discovery within maritime trajectories can lead to significant benefits, such as identifying optimal docking locations, which is vital given the high costs associated with boat docking. This can significantly impact overall trajectory expenses [12], allowing maritime operators to strategically allocate resources and minimize unnecessary expenses [5]. Moreover, ZOI discovery contributes to optimizing trajectories, enhancing efficiency, and reducing fuel consumption. Adherence to site-specific environmental and security regulations, such as constraints on pollutant emissions (e.g., CO_2, NO_2, SO_2) and the designation of protected areas, further underscores the importance of ZOI knowledge.

Given the significant role that ZOIs play in maritime trajectories, it is crucial to develop a methodology for their automatic discovery and labeling. This paves the way for the semantic annotation of trajectories, leading to a better understanding of maritime activities and improvements in maritime transport efficiency through a data-driven approach.

In this paper, we describe a framework for the discovery, description, and annotation of ZOIs within maritime trajectories. This framework is developed as part of the TNTM project [10] which aims to optimize the transportation system and address environmental concerns in line with the International Maritime Organization's strategy for reducing Greenhouse Gas emissions from ships [8]. The methodology behind the framework is data-driven, relying on raw trajectory data mining and contextual metadata for ZOI labeling.

The rest of this paper is organized as follows: In Section 2, we outline the research issues and define formally a ZOI. In Section 3, we review related works in the field. Section 4 offers a detailed description of the framework, emphasizing its structure and components. We discuss a simple use case to demonstrate the main benefits of our framework in Section 5. Finally, Section 6 concludes the paper by highlighting ongoing work and future research directions.

2. Problem statement

Formally, a ZOI can be defined and modelled as an extension of the Open Geospatial Consortium (OGC) features data model [11], representing spatial objects that are significant for various stakeholders. In the maritime domain, ZOIs include not only areas within the sea but also coastal objects such as ports, oil and fuel stocking tanks, and other related infrastructures. A ZOI comprises two essential dimensions: the physical or spatial aspect and the logical or contextual representation. The first dimension involves defining and delineating the geographical area of interest, capturing its spatial boundaries. The contextual aspect provides descriptive insight into the activities corresponding to the ZOI. Hence, there is a necessity for a solution that encompasses both the physical and logical representations of ZOIs, ensuring an effective approach to ZOI discovery and annotation.

Figure 1 illustrates the ZOI data model, an extension of the OGC features data model, defining a ZOI and highlighting its physical and contextual dimensions.

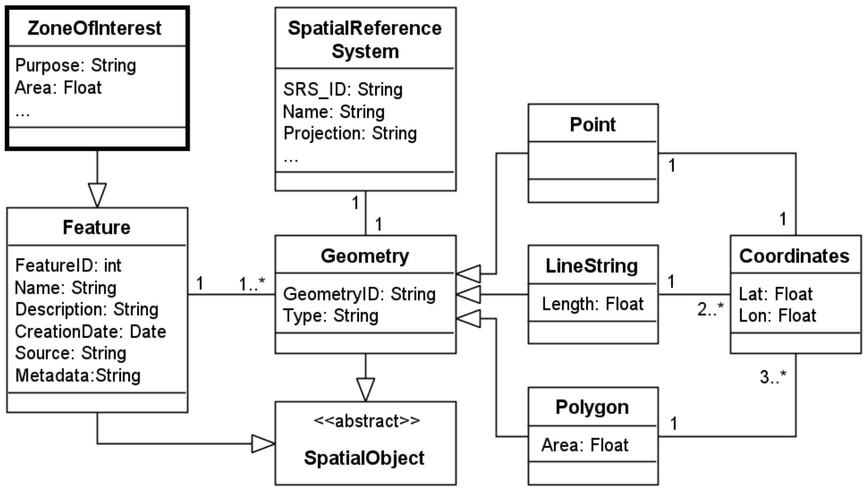


Fig. 1. ZOI data model extended from OGC features model

Framework Overview To meet the spatial and contextual dimensions of a ZOI criteria, we propose a threefold framework to process and transform raw maritime spatio-temporal trajectory data, collected from vessels and/or containers, into ZOI-wise trajectory data where every ZOI is annotated. In the first step we identify crucial stop points from raw trajectory data, which delineate primitive boundaries shapes of ZOIs. These shapes are then used to retrieve surface metadata from Volunteered Geographic Information (VGI) sources, such as OpenStreetMap[1] (OSM). OSM is particularly valuable due to its extensive and continuously updated geographic database. Finally, we use this contextual data to refine the ZOIs shapes and classify the activities they represent. The whole process is detailed in Section 4.

Challenges To develop this framework effectively, we face three primary challenges categories:

1. **ZOI Discovery**:
   Locating ZOIs within spatio-temporal trajectory data and accurately constructing their shapes is complex due to the dynamic nature of maritime activities. Factors such as swell waves, port management tasks, dock activities, changes in vessel paths, and weather conditions contribute to this complexity.
   Challenge examples: Locating ZOIs within trajectory data ; Delineating precise ZOI boundaries due to their dynamic nature
2. **ZOI Annotation**:
   Classifying the types of activities within ZOIs is another challenge. Indeed, classification requires a basis, and to our knowledge, no public ZOI types database exists. Thus, we need to construct a database of annotated ZOIs to unlock benefits like optimized docking costs, improved trajectories, reduced shipping emissions, and enhanced logistical ecosystems.

[1] https://www.openstreetmap.org

Challenge examples: Classification based on semi-structured data ; Utilizing unlabelled data for training ; Identifying possible ZOI types and classes

3. **Contextual Data Quality**:
Collecting contextual metadata from VGI sources, such as OpenStreetMap, presents quality issues [9], including lack of usable data, imprecise forms, missing tags, heterogeneity of tags, and 'expansionist' relationships. Preparing a tag dataset for ZOI classification based on their tags poses additional challenges, such as selecting representative zones and processing tags manually or semi-automatically.
Challenge examples: Inconsistencies in data representation and tagging ; Challenges in ensuring data accuracy and completeness ; Heterogeneity in data quality and tag distribution across different regions

3. Related work

Trajectory mining as discussed is widely applied in various domains, such as transportation, urban planning, social network analysis, environmental monitoring, etc. Within this broad context, the discovery and analysis of ZOIs have not gained yet significant attention in trajectory mining, despite their potential applications across various fields. This section delves into existing literature related to ZOI discovery and comparable studies that partially address this topic, discussing the gaps and challenges that our proposed framework aims to address.

To the best of our knowledge, the literature does not provide works specifically dedicated to the discovery and annotation of ZOIs. Most of the existing works focus on stops detection within trajectories, which can imply that a stop represents a significant point in the trajectory. However, stops alone do not necessarily indicate meaningful places or ZOIs, as they lack contextual semantics and do not specify the nature or type of these places. Despite the emphasis of these works on stops extraction within trajectories, we propose a comparative analysis to highlight the gaps targeted by this work. This framework not only identifies ZOIs but also precisely delineates their physical boundaries and annotates them with contextual information, specifically targeting the maritime domain where there is a lack of on-ground data to aid in context, classification, and annotation of ZOIs.

The SMoT model [1] enriches trajectories with semantic geographical information by extracting moves and stops using predefined geometries on the ground. This method efficiently identifies moves and stops, enriching trajectory data with semantic information. However, SMoT relies on predefined geometries, limiting its adaptability. An improved version, CB-SMoT [14], extracts additional stops not considered by SMoT as long as they do not necessary intersect with the predefined regions. While CB-SMoT enhances SMoT by identifying more stops, it remains dependent on predefined geometries.

Trajectory analysis using the Mobility Context Cube [13] integrates some contextual data, precisely points of interest opening hours, with extracted stops to classify real stops from fake ones using an SVM classifier. This approach effectively combines limited contextual data with trajectory analysis but does not address the precise spatial shape of ZOIs or extensive contextual dimension. This work primarily focuses on urban mobility contexts.

The framework proposed by Bisone [3] for extracting the operational behaviors of emergency vehicles using GPS data to optimize fleet management and route planning. It focuses on optimizing operational behaviors and utilizes GPS data to identify vehicles activities based on operational contexts. However, this work relies on physical devices installed inside the emergency vehicles - such as handbrake detection for instance - to extract stops, more precise shapes and contextual identification of ZOIs. This highly impacts the framework adaptability and applicability to other cases or scenarios as it is specifically limited to emergency vehicles' data.

DJ-Cluster [15] extends the DBSCAN algorithm to extract significant locations or ZOIs based on trajectory data by introducing density-joinable clusters. This method is effective for identifying significant locations using density-based clustering but focuses solely on spatial aspects, neglecting temporal, precise ZOI shape and contextual dimensions.

ST-DBSCAN [2] considers both spatial and non-spatial aspects of data to extract stops within trajectories. It integrates non-spatial dimensions with spatial data, providing a more comprehensive approach to stop extraction. Despite this, it lacks focus on precise ZOI shape delineation and lacks contextual data integration for annotation. Similarly, T-DBSCAN [4] combines temporal and spatial dimensions to mine stops from trajectories. While effective in combining temporal and spatial data for stop extraction, it does not address precise ZOI shape delineation or contextual information integration.

This Paper presents an extended and refined version of our previous work on trajectory mining and ZOIs discovery [7], introducing a matured and enhanced framework that addresses several gaps in the related works. Table 1 offers a summarized view of the related works drawbacks, and presents how our framework benefits from each of the related works. Our framework addresses the gaps identified in the existing literature by providing comprehensive ZOI discovery, precise spatial shape determination, contextual aspect identification, and cross-domain adaptability, without relying on information from predefined ZOIs. This integrated approach ensures an effective solution for ZOI detection and classification within spatio-temporal trajectory data. While the description of our framework is tailored for maritime data in this paper, the methodology is broadly applicable to various transportation domains, including vehicular and pedestrian trajectories.

4. Framework Proposal

In this section, we present an extended version of our previous framework [7], designed to identify ZOIs within trajectory data. While this work mainly focuses on the maritime domain, the framework is adaptable across different domains. Our approach includes several steps to detect, outline, and classify ZOIs efficiently. By combining raw trajectory data with contextual on-ground information and industry expertise, we aim to improve the accuracy and precision of ZOI detection and characterization. Our framework unfolds in three phases:

1. **Extraction of starts and stops from trajectory data.** We use a custom density-based algorithm, that we name STC-DBSCAN (see 4.1), to extract stop points within raw vessels and/or containers trajectories. Then, we conduct a second clustering layer to identify the most important stops areas, which form the initial shape of the ZOIs.

Table 1. Benefits and limitations of related works

Work	Stop Extraction	Precise ZOI Spatial Shape	Contextual Aspect	Cross-Domain	Non-Predefined ZOI Info
SMoT [1]	✓	✓		✓	
CB-SMoT [14]	✓			✓	
Mobility Context Cube [13]	✓			✓	
Emergency Vehicles Activity Identification [3]	✓	✓	✓		✓
DJ-Cluster [15]	✓			✓	✓
ST-DBSCAN [2]	✓			✓	✓
T-DBSCAN [4]	✓			✓	✓
This Framework	✓	✓	✓	✓	✓

2. **Context retrieving and ZOI shape refinement.** We leverage the constructed shapes to retrieve on-ground contextual data from open geographical base, particularly, OpenStreetMap (OSM). More specifically, we extract textual metadata (OSM tags) associated with all the objects located within or intersect with the initial ZOIs shapes. We then refine and construct a more realistic shape using the OSM objects' polygons.
3. **ZOI description and annotation.** We use the textual content of OSM tags to create a ZOI classification model. We then run all the extracted ZOI into this model to describe and annotate them.

The overall framework is depicted in Figure 2. For clarity, a simplified view of the framework is provided. Each step, indicated by a colored rectangle, is detailed in its corresponding subsection. The threefold splitting ensures a comprehensive approach to ZOI detection, refinement, and annotation, leveraging trajectory data, on-ground contextual data, and domain-specific knowledge.

4.1. Step 1: start/stop detection

The initial step of the framework involves detecting start and stop points from raw trajectory data, followed by a clustering process to extract the initial shapes of ZOIs. This step employs two algorithms, each designed to achieve one of the aforementioned objectives. The process of Step 1 is depicted in Figure 3.

Start & Stop detection To extract stops positions from raw trajectory data, we leverage the STC-DBSCAN algorithm, which we previously introduced in [7]. As described in Algorithm 1, STC-DBSCAN extends the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method to identify potential stop points based on density. It subsequently refines these points by evaluating their density levels and temporal characteristics to distinguish between genuine stops and slowdown areas.

DBSCAN is a well-known clustering algorithm used to identify clusters of varying shapes and sizes in spatial data. It regroups points that are closely packed together while

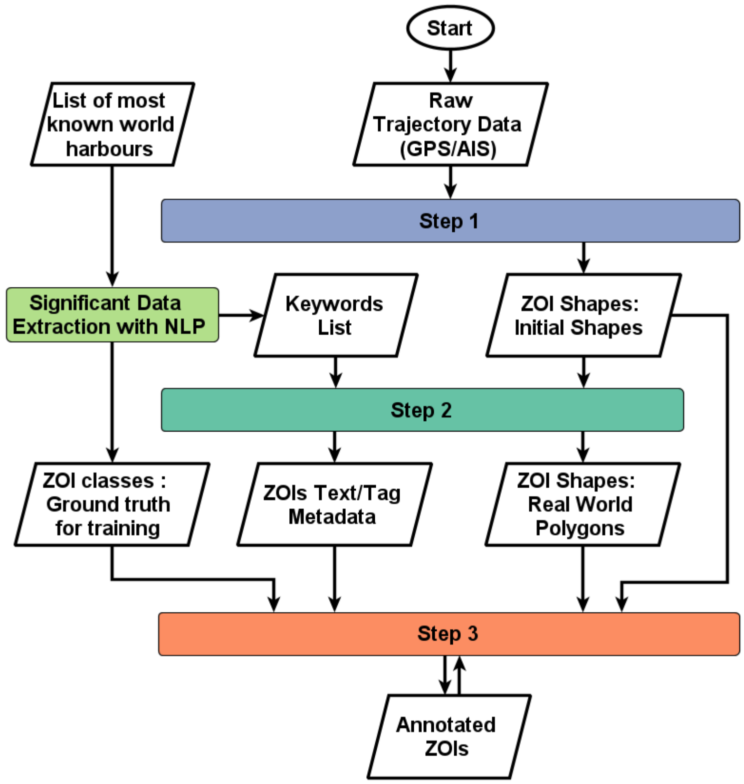


Fig. 2. ZOI Discovery and Annotation Framework

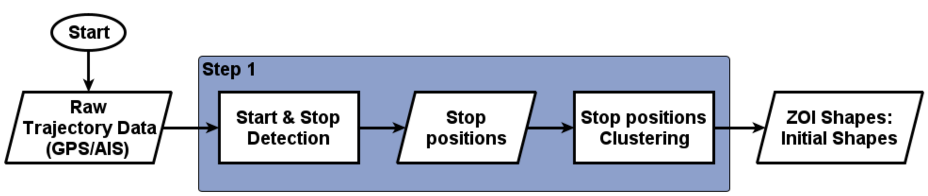


Fig. 3. Framework Step 1 : Start & stop detection

marking points that lie alone in low-density regions as outliers. By examining the density of data points in a given neighborhood, DBSCAN can effectively detect clusters and distinguish noise, making it suitable for applications where the data contains clusters of arbitrary shapes and noise.

Although the STC-DBSCAN algorithm has not yet undergone formal testing, we have initiated the development of a benchmark. Our objective is to compare our algorithm's

Algorithm 1: Brief representation of the **S**patio**T**emporal **C**lustering - DBSCAN Algorithm (STC-DBSCAN)

```
Data: T : trajectory points, t1 : spatial threshold, t2 : temporal threshold
Result: ST : Trips, S : Stops
if consecutive points are within the threshold distance then
    Delineate stop zone beginning;
else
    Delineate the end of stop zone if already initiated, then start a trip
end
for ( each stop zone z ) {
    if z represents high density (BboxDiagDistance < t1) && (cumulatedTime in z > t2)
      then
        Validate the stop zone
    else
        Represents a slowdown area in the current trip
    end
}
```

performance with that of several well-known algorithms in the context of start-stop extraction from raw trajectory data.

Stop positions clustering After extracting stops from all trajectories, we proceed to a clustering phase to identify areas of high stop density. Using the extracted stops as input, we apply the DBSCAN algorithm to cluster regions with concentrated stop occurrences, as the DBSCAN is suitable to construct clusters of arbitrary shapes. This step is crucial because individual stops within a trajectory do not necessarily indicate the existence of a ZOI. The reasons for individual stops can vary widely and may not always signify areas of importance. However, by identifying clusters with high stop density, we can more reliably locate areas likely to represent ZOIs. This allows us to establish the initial shape of the ZOI.

4.2. Step 2: Context retrieving and ZOI shape refining

Through the first step of the framework, we have identified areas of significance within maritime trajectories. These areas may be offshore (vessels paths or stops for instance), or on the coast (docking, container loading or unloading for instance). These zones hold significance, yet lack clear identification or identity. While their importance is evident, they require human intervention to ascertain their precise nature each time they are encountered. In order to enhance the discovery process and automate it, as well as to annotate our ZOIs, we retrieve meaningful contextual data that can help in ZOI identification.

Furthermore, the stop position clustering algorithm from step 1 provides coarse ZOI shapes. These shapes are determined by the raw trajectories dataset and hold two main drawbacks : 1) the shapes are dependent of several maritime factors such as changes in vessels paths or weather conditions, and 2) the shapes are not geographically refined as they do not detect nor encompass human buildings such as ports or docks.

The second step of our framework aims to achieve such goals through VGI context retrieving. The detail of this step is presented in Figure 4.

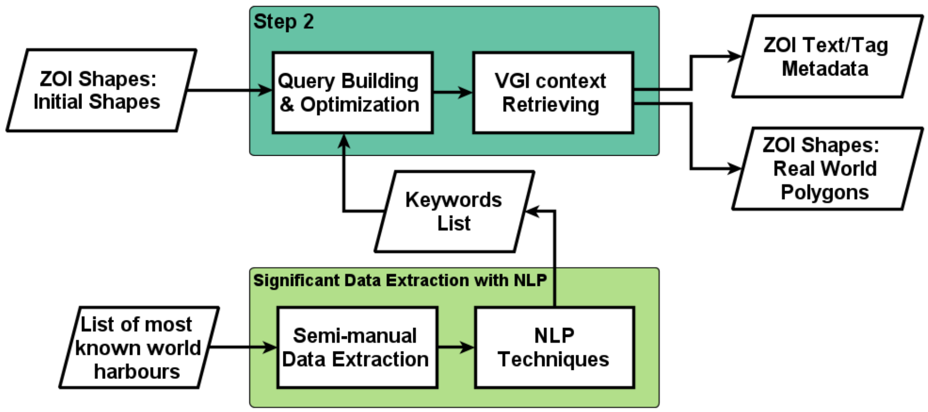


Fig. 4. Framework Step 2 : ZOI shape refining

The second step is divided into two main processes. The first process, named *"step 2"* in Figure 4 consists in implementing on-ground information retrieval over the initial ZOI shapes extracted in step 1. This process runs queries on VGI resources to extract contextual tags metadata, and precise polygons shapes. The second process, termed *"Significant data extraction with NLP"* involves constructing a list of keywords to facilitate VGI querying and, subsequently, in sub-section 4.3, to establish a set of ZOI types or classes. Each process is described in detail below.

VGI Context Retrieving To extract contextual metadata from the initial ZOIs, we run customized queries over the OSM database using the Overpass API tool[2]. The API retrieves relevant on-ground textual and geographical information (called tags), providing valuable insights about the ZOIs. Additionally, we use this metadata to construct refined ZOIs shapes using the actual OSM objects presents within these areas.

This process faces significant challenges due to data quality issues within the OSM database as discussed in section 2. The structure of OSM data, consisting of *nodes, ways*, and *relations*, does not consistently provide accurate logical representations of objects and their relationships. For instance, while attempting to retrieve objects present within a ZOI, the *relations* can expand extensively and retrieve irrelevant data and objects, posing a significant risk if integrated into a larger system or environment for further processing. We present an illustrative example in Figure 5. We consider a scenario where a small ZOI inadvertently retrieves data encompassing the entirety of the Arabian Sea. This occurrence illustrates the challenges of using such semi-structured data, especially the *relation* tags. Such anomalies highlight the need for careful consideration and mitigation strategies to address data quality issues effectively.

Significant Data Extraction with NLP To take up this challenge, we propose an adapted workflow that enhances the overall framework and specifically targets this phase and its issues. The main idea of this part is to build a list of keywords that specifically target maritime objects or infrastructures, and to use this list as a filtration basis aimed at restricting the scope of OSM queries research and their *relations* expansion.

[2] https://overpass-turbo.eu

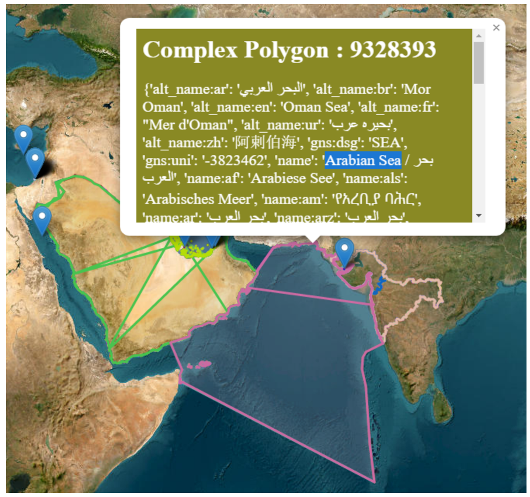


Fig. 5. Example of data quality issues

By incorporating specific keywords, we establish criteria that ensure only relevant objects are retrieved using the OSM Overpass engine. This approach mitigates the risk of extensive expansion caused by *relations* in the OSM data. By enforcing a requirement for objects to contain at least one of the specified keywords, we streamline the query process to prioritize the retrieval of useful data and mitigate quality issues. For instance, missing tags in some objects, variations in tag presence across objects, and data heterogeneity arising from multiple languages and regional verification practices.

To address these challenges, we propose a solution focused on the semi-automatic generation of a keywords database. In an initial version of our framework, we suggest selecting a representative sample of the largest and most active ports from the OSM dataset. By conducting data profiling on these ports, we aim to extract statistics on tag usage to identify consistent keywords indicative of port-related features. Additionally, while industrial keywords are essential, their indiscriminate usage may lead to inaccurate results due to the widespread presence of industrial zones worldwide. Hence, careful selection and refinement of keywords, along with tailored queries designed for OSM, are crucial for effectively leveraging the dataset for ZOI discovery and annotation.

4.3. Step 3: ZOI annotation

The final part of our framework involves three main processes: the reconstruction, classification and annotation of the ZOIs. The detail of this step is presented in Figure 6.

ZOI shape reconstruction Through the filtration process from step 2, we identify relevant on-ground objects inside a ZOI and extract their shapes. Then we merge the initial ZOI shape with the shapes of these objects and create a refined and final shape that accurately represents the ZOI.

With this final process we construct final ZOIs shapes that hold maritime and human meanings. Infrastructures such as port, docks, terminals, canals, are added to the initials ZOIs shapes, allowing a more precise contextual annotation of each ZOI.

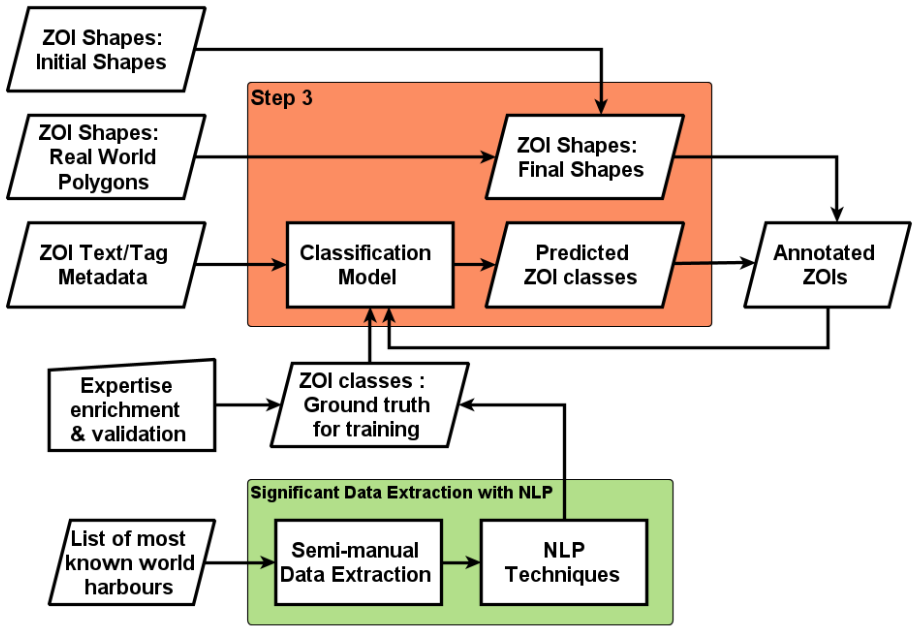


Fig. 6. Framework Step 3 : ZOI annotation

ZOI Classification The second process of this step consists in classifying each ZOI. This classification process is done using a list of ZOI classes which is built from two main sources. First, using the tags database from step 2, we conduct a clustering phase to identify potential ZOI classes. These classes are then validated and refined through expertise and business knowledge. In the TNTM french project, these classes will be validated by the main maritime transportation authority leading the project.

Once a list of maritime ZOI classes, or types, are defined, we plan to use large language models (LLMs) like BERT, embedding techniques such as Word2Vec or TF-IDF, and various text classification methods to classify each ZOI. These methods will leverage the tags and data extracted from step 2 to predict the class of each ZOI, ensuring a comprehensive and accurate annotation process.

ZOI Annotation At the conclusion of our framework, we merge the final shape of our ZOI with its predicted class. This integration results in an annotated ZOI, offering both spatial and contextual representation. Additionally, this annotation serves as valuable feedback for enhancing the classification model iteratively and progressively.

5. Experiments

The experiments for this study are conducted using a set of maritime trajectory data from the TNTM project. These data hold vessels and/or containers positions. Currently, our framework effectively detects and delineates ZOIs, and extract OSM tags and shapes. This section presents a simple use case with experimental results, focusing on demonstrating the capabilities and outcomes of our framework through some ZOIs identification examples.

5.1. Initial ZOI Shape Construction

The first part of our framework involves detecting and extracting stops from trajectory data and constructing initial ZOI shapes. Figure 7 illustrates this process with the STC-DBSCAN algorithm. With a focus on the *Jawaharlal Nehru* port, India, we identify significant stop points and delineate initial ZOI shapes from several raw trajectories. Sub-figure 7a, represents different stops identified by the STC-DBSCAN algorithm. Rounded points represent different vessels and/or containers positions, while star-shaped points represent the detected stops. Sub-figure 7b represents the construction of the initial ZOI shape using the DBSCAN algorithm over the detected stops. As we discussed in previous sections, this shape is raw and do not encompass human activities and infrastructure. The next steps consists in refining this ZOI shape to include the totality of the *Jawaharlal Nehru* port infrastructures.

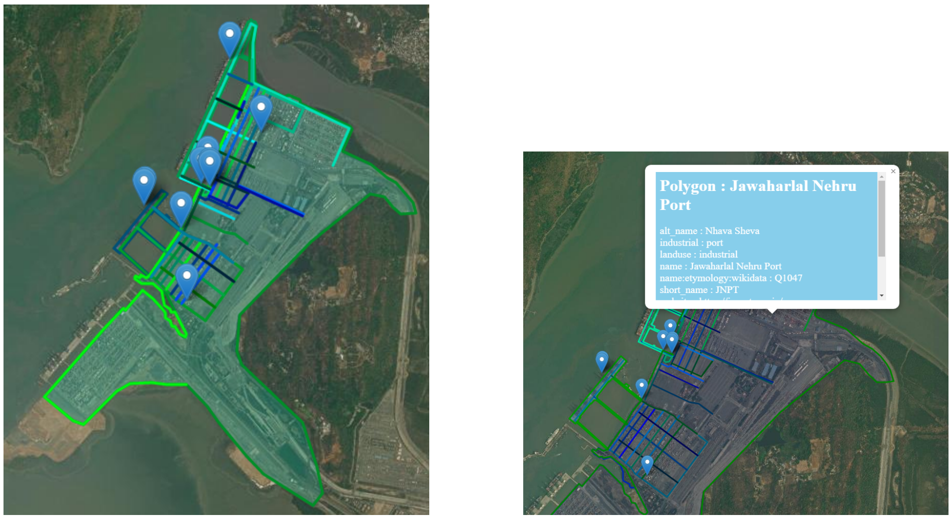
(a) Stops detected with STC-DBSCAN algorithm

(b) Initial ZOI shapes extracted with DBSCAN algorithm

Fig. 7. Stops detection and initial ZOI construction: example of the *Jawaharlal Nehru* port

5.2. ZOI Shape Refinement and Context Retrieval

The second part of our framework refines the initial ZOI shapes using contextual data retrieved from OSM. We extract textual metadata associated with the objects located or intersect with the initial ZOI shapes and reconstruct more precise and realistic ZOI shapes. Figure 8 illustrates two perspectives of the refined shape of the *Jawaharlal Nehru* port ZOI, along with its retrieved description. On the left, sub-figure 8a displays the shapes on a satellite map. We see the retrieved objects shapes related to the initial ZOI, such as buildings, roads, as well as the docks where the ship are anchored. The pin-shaped objects represent various ZOI-related OSM node objects, as also correspond to the defined stops clusters within the ZOI. These objects contribute to constituting the final ZOI shape. In sub-figure 8b, we see the descriptive texts of the ZOI. Here, for example, we display the textual information from the polygon presenting main shape of the *Jawaharlal Nehru* port, providing further insights about the ZOI characteristics.



(a) Transportation Map View

(b) OSM Tags descriptions

Fig. 8. ZOI shape refining using VGI Context Retrieving: example of the *Jawaharlal Nehru* port

5.3. ZOI Annotation

The final part of our framework involves the classification and annotation of the ZOIs. By utilizing the textual content of OSM tags, we feed a trained ZOI classification model to classify and annotate the ZOIs. Sub-figure 8b depicts the retrieved contextual information that aids in the classification and annotation of these ZOIs. The classification model is currently under development, and we discuss this topic in the following subsection.

5.4. Discussion

The objective of this paper is to propose and demonstrate a framework for the discovery and annotation of ZOIs. Although the benchmarking of the stop extraction algorithm and the text classification model for ZOI classification are currently under development, these are not the primary objectives of this paper. The main challenge in conducting an extensive practical comparison study of the proposed framework is the lack of existing frameworks in the literature that cover the dimensions of ZOI discovery and annotation as comprehensively as ours. While individual components, such as the stop extraction algorithm and the text classification model, can be benchmarked and compared to existing literature, this paper focuses on demonstrating the overall framework for ZOI discovery and annotation.

The elements presented in this section highlight examples of the outcomes of our framework in identifying and refining ZOIs within maritime trajectories. The precise shapes and contextual descriptions of ZOIs are crucial for realizing the benefits mentioned previously. Future work will involve benchmarking the stop extraction algorithm and the text classification model to further validate and enhance the proposed framework.

6. Conclusion

Analysis/mining ship trajectories plays a crucial role for various purposes such as preventing vessel collision or optimizing vessel energy consumption, to cite a few.

This paper describes a machine-learning based approach and framework for the discovery, refinement and annotation of Zones of Interest (*aka* ZOI) in a maritime trajectory. Our approach integrates trajectory analysis with contextual data and business insights, providing accurate and precise ZOI characterization. The framework consists mainly of a three-step process: (1) extraction of stops to devise the initial ZOI shape, (2) shape refinement by means of keyword-based contextual data retrieval, and (3) ZOIs' classification and annotation.

We believe this framework may bring a useful piece in the maritime trajectory mining landscape, but also, in other mobility application domains. As a future work, we envision benchmarking of the framework components and its deployment with real world data vessel trajectories, supplied by the TNTM consortium.

References

1. Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., Vaisman, A.: A model for enriching trajectories with semantic geographical information. In: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems. GIS '07, Association for Computing Machinery, New York, NY, USA (2007), https://doi.org/10.1145/1341012.1341041
2. Birant, D., Kut, A.: St-dbscan: An algorithm for clustering spatial–temporal data. Data Knowledge Engineering 60(1), 208–221 (2007), https://www.sciencedirect.com/science/article/pii/S0169023X06000218, intelligent Data Mining
3. Bisone, Frédérick, Étienne, Laurent, Devogele, Thomas: Modélisation et extraction de la sémantique des trajectoires à partir de données multicapteurs. Rev. Int. Geomat. 28(4), 461–483 (2018), https://doi.org/10.3166/rig.2018.00065
4. Chen, W., Ji, M., Wang, J.: T-dbscan: A spatiotemporal density clustering for gps trajectory segmentation. International Journal of Online and Biomedical Engineering (iJOE) 10(6), pp. 19–24 (Oct 2014), https://online-journals.org/index.php/i-joe/article/view/3881
5. Clark, X., Dollar, D., Micco, A.: Port efficiency, maritime transport costs, and bilateral trade. Journal of Development Economics 75(2), 417–450 (2004), https://www.sciencedirect.com/science/article/pii/S0304387804000689, 15th Inter American Seminar on Economics
6. Fratila (Adam), A., Gavril (Moldovan), I.A., Nita, S.C., Hrebenciuc, A.: The importance of maritime transport for economic growth in the european union: A panel data analysis. Sustainability 13(14) (2021), https://www.mdpi.com/2071-1050/13/14/7961

7. Ghannou, O.: Automatic discovery of zones of interests with maritime trajectory mining. In: Abelló, A., Vassiliadis, P., Romero, O., Wrembel, R., Bugiotti, F., Gamper, J., Vargas Solar, G., Zumpano, E. (eds.) New Trends in Database and Information Systems. pp. 684–692. Springer Nature Switzerland, Cham (2023)
8. International Maritime Organization: 2023 IMO Strategy on Reduction of GHG Emissions from Ships. https://www.imo.org/en/OurWork/Environment/Pages/2023-IMO-Strategy-on-Reduction-of-GHG-Emissions-from-Ships.aspx (2023), accessed on July 19, 2024
9. Kaur, J., Singh, J., Sehra, S.S., Rai, H.S.: Systematic literature review of data quality within openstreetmap. In: 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS). pp. 177–182 (2017)
10. Pôle Mer Méditerranée: Transformation Numérique du Transport Maritime. https://polemermediterranee.com/domaines-dactions-strategiques/transformation-numerique-du-transport-maritime/ (2022), accessed on July 20, 2024
11. W3C Group, OGC Document Number: OGC 15-107: Open Geospatial Consortium (OGC) features data model. https://www.w3.org/TR/sdw-bp/#spatial-things-features-and-geometry (2023), accessed on July 22, 2024
12. Wilmsmeier, G., Hoffmann, J., Sanchez, R.J.: The impact of port characteristics on international maritime transport costs. Research in Transportation Economics 16, 117–140 (2006), https://www.sciencedirect.com/science/article/pii/S0739885906160060, port Economics
13. Wu, T., Shen, H., Qin, J., Xiang, L.: Extracting stops from spatio-temporal trajectories within dynamic contextual features. Sustainability 13(2) (2021), https://www.mdpi.com/2071-1050/13/2/690
14. Xiu-Li, Z., Wei-Xiang, X.: A clustering-based approach for discovering interesting places in a single trajectory. In: 2009 Second International Conference on Intelligent Computation Technology and Automation. vol. 3, pp. 429–432 (2009)
15. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering personally meaningful places: An interactive clustering approach. ACM Trans. Inf. Syst. 25(3), 12–es (jul 2007), https://doi.org/10.1145/1247715.1247718

Omar Boucelma is a full Professor in Computer Science at Aix-Marseille University. He has a background in statistics (data analysis) and data management and has been conducting R&D projects in core databases and geospatial systems and applications. As part of his international research and educational activities, he was a visiting research fellow at CU Boulder, NUS Singapore and UNICAMP (Brazil). He has been involved in several French and (DARPA) American research projects, involving academia and/or industry. His recent works relate to geospatial entity matching, data streams and mobility data analytics.

Etienne Thuillier received a DUT (Technology degree) in computer sciences from the IUT of Lannion, France, in 2010 and the title of engineer in computer sciences (M.S. degree) at Belfort-Montbéliard University of Technology (UTBM), France, in 2013. He then obtained the Ph.D. degree in computer sciences at the FEMTO-ST/DISC OMNI team, at the UTBM. He is currently assistant professor at the Aix-Marseille University (AMU) at the LIS lab. In 2014 he was team a leader as part of the ComplexCity project between the UTBM and the Institute of Smart City, at Shanghai University Sino-European University of Technology (UTSEUS), Shanghai, China. His research interest includes

Intelligent Transportation Services (ITS) technologies, and more generally human and urban mobility.

Omar Ghannou is a PhD student at Aix-Marseille University with a background in machine learning. He has experience in domain adaptation, federated learning, and other diverse areas of the field, along with previous experience in the industry. His current research focuses on geospatio-temporal data mining.

Received: February 29, 2024; Accepted: August 06, 2024.