The first part of this book provides the motivating points and background for mobility forecasting supported by trajectory-oriented analytics. It presents specific problems and challenges in the Aviation (Air Traffic Management (ATM)) and the Maritime domains, clarifies operational concerns and objectives in both domains and explains domain-specific terminology. It presents challenging cases which motivate technology presented in subsequent chapters of this book and which drive evaluation and validation of solutions presented. Equally important to the above is the presentation of the data sources exploited, the big data challenges ahead in both domains and of course, the requirements from technologies presented in subsequent parts of the book.

Mobility Data: A Perspective from the Maritime Domain

Cyril Ray, Anne-Laure Jousselme, Clément Iphar, Maximilian Zocholl, Elena Camossi, Richard Dréo

Abstract This chapter[1] overviews maritime operational situations and underlying challenges that the automated processing of maritime mobility data would support with the detection of threats and abnormal activities. The maritime use cases and scenarios are geared on fishing activities monitoring, aligning with the European Union Maritime Security Strategy. Six scenarios falling under three use cases are presented together with maritime situational indicators expressing users' needs when conducting operational tasks. This chapter also presents relevant data sources to be exploited for operational purposes in the maritime domain, and discusses the related big data challenges to be addressed by algorithmic solutions. An integrated dataset of heterogeneous sources for maritime surveillance is finally described, gathering 13 sources. This chapter concludes on the generation of specific datasets to be used for algorithms evaluation and comparison purposes.

Cyril Ray and Richard Dréo
Arts et Metiers Institute of Technology, Ecole Navale, IRENav, F-29240 Brest, France, e-mail: `\protect\T1\textbraceleftcyril.ray,richard.dreo\protect\T1\textbraceright@ecole-navale.fr`

Anne-Laure Jousselme, Clément Iphar, Maximilian Zocholl and Elena Camossi,
NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy, e-mail: `\protect\T1\textbraceleftanne-laure.jousselme,clement.iphar,maximilian.zocholl,elena.camossi\protect\T1\textbraceright@cmre.nato.int`

[1] The material presented in this chapter is a compilation of excerpts from [7, 8, 14, 15].

Extended Abstract

Motivating points

Ensuring *security* and *control* of fishing activities is one of the most important aspect of the European Union Maritime Security Strategy () - Action Plan, which defines several strategic interests for the European Union and the Member States. Europe is the world's biggest market for seafood and the aim of the is to promote better international governance across the world's seas and oceans to keep them clean, safe and secure. Since fishing is an activity that exploits common natural resources, it needs to be regulated to safeguard fair access, sustainability and profitability for all.

Creating new capabilities for monitoring and supporting quick actions against Illegal, Unreported and Unregulated (IUU) fishing, human trafficking, accidents or collisions at sea is a major step for the successful implementation of the . Effective involves extracting relevant contextual information (for instance maritime routes or loitering areas) but also monitoring the real time maritime traffic. The use of a set of various sensors mixing cooperative self-identification systems such as the AIS and non-cooperative systems such as coastal radars or satellite imagery provides complementary and redundancy in information, as necessary to overcome the quite common spoofing of AIS signals and increase the clarity and the accuracy of the maritime situation. In many cases, intelligence information can also be helpful in refining and guiding the search in the huge amount of data to be processed, filtered and analysed, as well as representing the contextual information for some Maritime Situation Awareness (MSA) problems.

The project aimed to develop novel methods for threat and abnormal activity detection in very large amount of moving entities. Both training and testing of maritime anomaly detection solutions require high quality datasets, with features challenging the specific research objectives. Effective MSA requires to take into account a large number of different data sources offering divers data types. Moreover, the applicability of innovative mobility data solutions requires the handling of big data and the capability to respond to the connected challenges, *i.e.volume*, *velocity*, *variety* and *veracity*. To motivate and support the development of vessel movement analytics algorithms within *datAcron* and to ensure their operational relevance, we developed six scenarios organised in three use cases. The scenarios explain the operational challenges linked to the monitoring of fishing activities in European waters, while relevant Maritime Situational Indicator (s) express users' needs for appropriate awareness of the situation, targeting expected outcomes for algorithms.

Abstract of the approach

This chapter overviews maritime operational situations and underlying challenges that the automated processing of maritime mobility data would support with the detection of threats and abnormal activities. The maritime use cases and scenarios are geared on fishing activities monitoring, aligning with the European Union Maritime Security Strategy. Indeed, fishing activity monitoring is a complex maritime surveillance mission that encompasses several maritime risks and environmental issues such as environmental destruction and degradation but also maritime accidents, fishing and trafficking problems. In particular, fishing is a global threat to the marine environment and honest fishermen alike, whose global cost is estimated in about 10 Billion Euros per year. The European Union, in collaboration with international organisations, is committed to fighting fishing worldwide. Besides the detection of fishing activities, safety is another core issue of the . Fishing vessels may fish in areas with dense traffic, like traffic lanes and waiting areas, and to keep the fishing place hidden from others, they sometimes intentionally switch off their AIS device while fishing, endangering themselves and the surrounding traffic.

Therefore, preserving the maritime environment from illegal fishing and ensuring fishing safety requires live processing and prediction of fishing vessel trajectories, identifying, thanks to the detection of MSIs, movement patterns (*e.g. close-encounters*, *change in speed* and *course*), detecting vessel activities (*e.g. fishing*, *loitering*, *tugging*, *rendez-vous*), forecasting potential collisions between surrounding ships within a typical time scale of 5 to 15 minutes.

Six scenarios falling under three use cases have been developed in collaboration with operational experts to capture domain requirements. They are presented together with maritime situational indicators expressing users' needs when conducting operational tasks. The scenarios highlight the needs for live tracking of fishing vessels and surrounding traffic, as well as of contextually enhanced offline data analytics, including for instance, cluster and spatial analysis together with motion pattern detection. These scenarios have been elaborated in order to (1) stress trajectory and event detection, prediction and visualisation algorithms against Big Data challenges in terms of *velocity*, *veracity*, *variety* and *volume*, and (2) provide operational relevance to their future use. Each scenario of the three use cases focuses on *secured fishing*, *maritime sustainable development* or *findexmaritime security* and highlights different users' **goals** and possible **actions**, as well as **information needs** expressed in terms of Maritime Situational Indicators.

This chapter also presents relevant data sources to be exploited for operational purposes in the maritime domain, and discusses the related Big Data challenges to be addressed by algorithmic solutions. requires processing in real-time a high volume of information of different nature (numerical, geographic, natural language statements, objective or subjective assessments, . . .), originating from a variety of sources (like sensors and humans, respectively hard and soft sources), with a lack of veracity (*i.e.* being uncertain, imprecise, vague, ambiguous, incomplete, conflicting, incorrect, etc). The algorithms to be designed in support to should cope with these Big Data challenges and this ability should be reflected in the quality of the results provided.

An integrated dataset of heterogeneous sources for maritime surveillance [15] is finally described, gathering 13 sources. We identified over forty data sources, classified within sixteen categories. Based on this preliminary study, a representative heterogeneous maritime dataset was built to support the developments of algorithms and visualisation features. The data description provides details about type, source, originator, file format (with , ,), spatial and temporal extent, size, and approximate average stream rate (with *msg*, messages; *obs*, observations). The maritime reference dataset relies on the most widely used maritime reporting system, AIS. The AIS is one of the electronic systems that enable ships to broadcast their position and nominative information via radio communications. In order to understand maritime activities and their impact on the environment, spatially and temporally aligned maritime data capturing additional features to ships' kinematic from complementary data sources (environmental, contextual, geographical, . . .) is of great interest. The dataset described in this chapter contains ship information collected through the AIS, prepared together with correlated contextual data, spatially and temporally aligned, characterising the vessels, the area where they navigate and the situation at sea. The dataset contains four categories of data: navigation data, vessel-oriented data, geographic data, and environmental data. It covers a time span of six months, from October 1st, 2015 to March 31st, 2016 and provides ship positions over the Celtic sea, the North Atlantic Ocean, the English Channel, and the Bay of Biscay (France). In addition a stream of data covering European coast has been considered to stress algorithms under higher volume and velocity.

This chapter concludes on the generation of specific datasets to be used for algorithms evaluation and comparison purposes. To support the evaluation of the algorithms, the reference dataset has been enriched with contextual historical information. Raw data have been processed to extract maritime patterns and clusters, specifically maritime routes, which facilitate the operators assessment. Additionally, a library of functions has been designed and implemented to systematically degrade and enrich the AIS batch dataset. The modifications applicable to each AIS field have been categorised along *kinematic*, *coverage* and *spoofing* dimensions. This library of functions provides a rich set of basic constructs to build different modification events or patterns in a non unique manner [5]. Having different ways to produce equivalent patterns or to provide a wide diversity in similar patterns is desirable to create realistic synthetic datasets. A semi-automatic process was designed to generate dedicated scenarios [23, 5] following several steps of data enrichment, data injection, resulting in a story matching the maritime surveillance challenges.

Main conclusions and results

Facing the huge *volume* of *various* information with high *velocity* which often lacks *veracity*, a system to automatically process both historical and timely information would greatly support the different maritime operators in monitoring and performing several types of analysis of situations at sea. This has been the aim of the three-year *Big Data Analytics for Time Critical Mobility Forecasting (datAcron)* project[2] that

[2] *datAcron* project website: `http://www.datacron-project.eu(accessedJanuary2020)`

run from January 2016 to December 2018 and whose main research objectives addressed the development of highly scalable methods for advancing (1) Spatio-temporal data integration and management solutions, (2) Real-time detection and forecasting accuracy of moving entities' trajectories, (3) Real-time recognition and prediction of important events concerning these entities, (4) General visual analytics infrastructure supporting all steps of the analysis through appropriate interactive visualisations, (5) Producing streaming data synopses at a high-rate of compression.

The operational scenarios and challenges together with the supporting datasets described in this chapter were aimed at supporting the design and development of Big Data analytics tools for maritime surveillance, such as the one developed in the project. Six scenarios corresponding to three use cases of Secure fishing, Sustainable development and Maritime security were proposed, while the collision avoidance scenario under the secure fishing use case was retained for experiments. A set of 28 Maritime Situational Indicators was proposed to both capture user information needs and provide targets to event detectors to be designed by the partners and integrated in the final prototype. The scenario was described with relevant s, for which big data analytics solutions were proposed and implemented. The algorithms thus designed have been exercised on an heterogeneous dataset gathering timely AIS data and other contextual data. The open heterogeneous maritime dataset has been made publicly available to enable other further maritime experimentation in realistic operational settings, challenging specifically the *variety* dimension of big data. The maritime prototype has been set-up following a human-centric approach involving maritime experts and tying together scenario definition, data preparation, visualisation and human evaluation. Finally, the evaluation of the maritime prototype has been conducted according to a methodology further described in [22].

Contributions

The contributions for this work are the following:

1. We developed of a methodology for articulating collaborative research in which the s play a pivotal role between the operational scenarios and the prototype design, capturing operators needs in a synthetic way on the one hand, and driving the development and integration of algorithms on the other hand.
2. We defined use cases and scenarios targeting several operational users performing maritime surveillance tasks. The scenarios described explicit user goals and information needs, together with actions to be taken once a suitable level of awareness was reached.
3. We proposed a list of s characterising low-level and generic user information needs and expressing output requirements for the design of algorithms.
4. We provided maritime data to the project partners as well as the research community. This publicly available integrated dataset from heterogeneous maritime data sources provides a benchmark against which algorithms can be compared.

5. We designed and implemented a methodology and a library to create, modify, enrich and degrade a set of data automatising the generation of scenario data.

Acknowledgements This work was supported by project *Big Data Analytics for Time Critical Mobility Forecasting* (datAcron), which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 687591. The authors would like to thank Christophe Claramunt, Melita Hadzagic, LCdr Eric Reardon, Cdr Mike Ilteris and Karna Bryan for their participation to the original ideas of the use cases and scenarios and cadets of the French Naval academy for their participation in the design of scenario datasets.

References

1. Andler, S., Fredin, M., Gustafsson, P.M., van Laere, J., Nilsson, M., Svenson, P.: SMARTracIn - a concept for spoof resistant tracking of vessel and detection of adverse intentions. In: Proceedings Volume 7305, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VIII; 73050G. Orlando, FL (2009)
2. Boudiere, E., Maisondieu, C., Ardhuin, F., Accensi, M., Pineau-Guillou, L., Lepesqueur, J.: A suitable metocean hindcast database for the design of marine energy converters. International Journal of Marine Energy **3-4**, e40–e52 (2013). DOI 10.1016/j.ijome.2013.11.010
3. Halpern, B.S., Frazier, M., Potapenko, J., Casey, K.S., Koenig, K., Longo, C., Lowndes, J.S., Rockwood, R.C., Selig, E.R., Selkoe, K.A., Walbridge, S.: Spatial and temporal changes in cumulative human impacts on the world's ocean. Nature communications **6** (2015)
4. Heumann, J.: Tips for writing good use cases. IBM - Transforming software and systems delivery (2008). White paper
5. Iphar, C., Jousselme, A.L., Ray, C.: Pseudo-synthetic datasets in support to maritime surveillance algorithms assessment. In: proceedings of the VERITA Workshop, 19ieme Journées Francophones Extraction et Gestion des Connaissances (EGC) 2019, Metz, France, January 21-25 (2019)
6. (ITU), I.T.U.: Technical characteristics for an automatic identification system using time division multiple access in the vhf maritime mobile frequency band (recommendation m.1371-5) (2014). URL itu.int/rec/R-REC-M.1371-5-201402-I
7. Jousselme, A.L., Ray, C., Camossi, E., Hadzagic, M., Claramunt, C., Bryan, K., Reardon, E., Ilteris, M.: Maritime use case description, H2020 datAcron d5.1 (2016)
8. Jousselme, A.L., Reardon, E., Hadzagic, M., Camossi, E., Bryan, K., Ilteris, M., Ray, C., Claramunt, C.: A maritime use case in support to collaborative research. In: M. Vespe, F. Mazzarella (eds.) Proc. of the Maritime Knowledge Discovery and Anomaly Detection (KDAD) Workshop. European Commission Joint Research Centre, European Commission Publication Office, Ispra, Varese (2016). DOI 10.2788/025881. URL https://bluehub.jrc.ec.europa.eu/static/KDAD/KDAD_Proceedings.pdf
9. Le Guyader, D., Ray, C., Brosset, D.: Defining fishing grounds variability with automatic identification system (AIS) data. In: 2nd International Workshop on Maritime Flows and Networks (WIMAKS'16) (2016)
10. Lee, J., South, A.B., Jennings, S.: Developing reliable, repeatable, and accessible methods to provide high-resolution estimates of fishing-effort distributions from vessel monitoring system (VMS) data. ICES Journal of Marine Science **67**(6), 1260–1271 (2010). DOI 10.1093/icesjms/fsq010
11. Mazzarella, F., Vespe, M., Damalas, D., Osio, G.: Discovering vessel activities at sea using AIS data: Mapping of fishing footprints. In: 17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, July 7-10, 2014, pp. 1–7 (2014). URL http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6916045

12. Natale, F., Gibin, M., Alessandrini, A., Vespe, M., Paulrud, A.: Mapping fishing effort through ais data. PloS one **10**(6), e0130746 (2015)
13. Ray, C., Camossi, E., Dréo, R., Jousselme, A.L., Iphar, C., Zocholl, M., Hadzagic, M.: Use case design and big data analytics evaluation for fishing monitoring. In: OCEANS 2019, Marseille, France, June 17-20 (2019)
14. Ray, C., Camossi, E., Jousselme, A.L., Hadzagic, M., Claramunt, C., Batty, E.: Maritime data preparation and curation, H2020 datAcron d5.2 (2016)
15. Ray, C., Dréo, R., Camossi, E., Jousselme, A.L., Iphar, C.: Heterogeneous integrated dataset for maritime intelligence, surveillance, and reconnaissance. Data in Brief **25** (2019). DOI 10.1016/j.dib.2019.104141
16. Ray, C., Grancher, A.: Integration of Nautical Charts in Maritime Dataset (2018). DOI 10.5281/zenodo.1182539. URL `https://doi.org/10.5281/zenodo.1182539`.
17. Ray, C., Iphar, C., Dréo, R., Kleynhans, W., Camossi, E., Jousselme, A.L., Zocholl, M., Batty, E., Roche, Q., Metzger, A.: Maritime data preparation and curation (final), H2020 datAcron d5.4 (2018)
18. Roy, J., Davenport, M.: Exploitation of maritime domain ontologies for anomaly detection and threat analysis. In: Proceedings of the 2010 International Waterside Security Conference (WSS). IEEE (2010). DOI 10.1109/WSSC.2010.5730278
19. Vespe, M., Gibin, M., Alessandrini, A., Natale, F., Mazzarella, F., Osio, G.C.: Mapping eu fishing activities using ship tracking data. Journal of Maps **12**(sup1), 520–525 (2016)
20. Vouros, G.A., Vlachou, A., Santipantakis, G.M., Doulkeridis, C., Pelekis, N., Georgiou, H.V., Theodoridis, Y., Patroumpas, K., Alevizos, E., Artikis, A., Claramunt, C., Ray, C., Scarlatti, D., Fuchs, G., Andrienko, G.L., Andrienko, N.V., Mock, M., Camossi, E., Jousselme, A.L., Garcia, J.M.C.: Big data analytics for time critical mobility forecasting: Recent progress and research challenges. In: M.H. Böhlen, R. Pichler, N. May, E. Rahm, S. Wu, K. Hose (eds.) Proceedings of the 21th International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018., pp. 612–623. OpenProceedings.org (2018). DOI 10.5441/002/edbt.2018.71. URL `https://doi.org/10.5441/002/edbt.2018.71`
21. Zocholl, M., Camossi, E., Jousselme, A.L., Ray, C.: Ontology-based Design of Experiments on Big Data Solutions. In: A. Khalili, M. Koutraki (eds.) Proceedings of the Posters and Demos Track of the 14th International Conference on Semantic Systems (SEMANTiCS 2018), Vienna, Austria, September 10-13, 2018., *CEUR Workshop Proceedings*, vol. 2198. CEUR-WS.org (2018). URL `http://ceur-ws.org/Vol-2198/paper_117.pdf`
22. Zocholl, M., Iphar, C., Camossi, E., de Rosa, F., Jousselme, A.L., Ray, C., Dréo, R.: User-centric assessment of maritime situation awareness solutions. In: OCEANS 2019, Marseille, France, June 17-20 (2019)
23. Zocholl, M., Iphar, C., Pitsikalis, M., Jousselme, A.L., Artikis, A., Ray, C.: Evaluation of maritime event detection against missing data. In: M. Piattini, P.R. da Cunha, I.G.R. de Guzmán, R. Pérez-Castillo (eds.) Quality of Information and Communications Technology - 12th International Conference, QUATIC 2019, Ciudad Real, Spain, September 11-13, 2019, Proceedings, *Communications in Computer and Information Science*, vol. 1010, pp. 275–288. Springer (2019). DOI 10.1007/978-3-030-29238-6_20. URL `https://doi.org/10.1007/978-3-030-29238-6_20`