

Data degradation variations for maritime situational indicator detection assessment

Clément Iphar¹, Anne-Laure Joussetme¹ and Cyril Ray²

¹ NATO STO CMRE, La Spezia, Italy

² Naval Academy Research Institute, Brest, France

{clement.iphar; anne-laure.joussetme}@cmre.nato.int
cyril.ray@ecole-navale.fr

Abstract. The identification and the prediction of maritime situational indicators (MSI), being of foremost importance in maritime scenarios assessment, must be assessed in the scope of degradation variation. In this respect, this paper presents a set of methods for maritime data controlled degradation enabling the analysis of a dataset with various veracity levels. A twofold methodology is proposed: for data controlled degradation and for MSI prediction assessment.

Keywords: Data degradation, Data veracity, Maritime Situational Indicator.

1 The V's of Big Data

Traditionally, four properties are associated with Big Data, which are called the four V's: Volume, Velocity, Variety and Veracity. The Volume is in relation with the amount of data to be handled, whereas the Velocity is effectiveness of gathering and processing. Being more applicative, enabling effectiveness and researches to be done more quickly, the Velocity of data exploitation is more important than the Volume.

The Variety property covers the fact that data in Big Data takes several forms [1], and most of those data sources are recent. The Veracity is a challenge as it is not linked to the quality of the data but with its relation to the world. It represents the fact for a datum to be truthful, *i.e.* to correctly depict the world in an expected way. Our prospective study concentrates on the application of variety and variety variations in the scope of AIS (Automatic Identification System) maritime positioning messages.

2 Degradation methods for veracity variations

The variations in veracity, in the case of AIS messages will be measured by the input data quality. AIS data is imperfect, as errors, falsifications and spoofing cases have been shown [2]. We are interested in means to degrade (enrich) data so that the dimension of veracity varies, and the data quality dimensions defined in [3] (completeness, accuracy, clarity, continuity and timeliness) can be evaluated. The purpose of degradation is to provide reference data based on a qualitative and quantitative data

analysis, with known quality levels so that various scenarios can be set to assess and validate MSI algorithms. The methods we propose in this work can be classified in four families: noise adjunction, data modification, data removal or data addition (Figure 1).

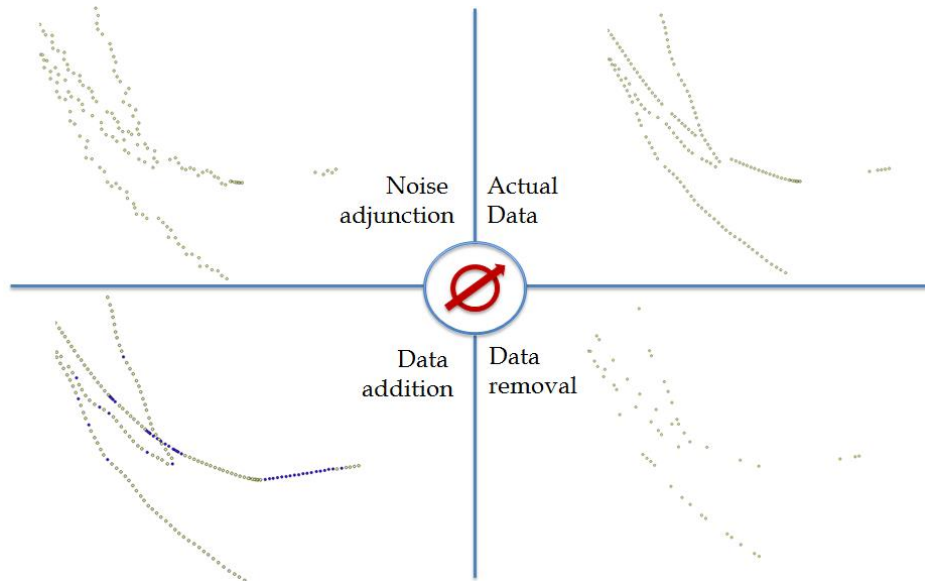


Figure 1: Controlled degradation of data quality

The application of those methods to datasets is intended to modify the inner quality of those datasets, and therefore to modify the outputs of analyses performed on the basis of their data. As the data fields within AIS messages are various, those selected methods will apply to selected data fields, and for instance, the noise adjunction, consisting in the blurring of the value by applying a Gaussian shift to it, can only be applied to physical values such as speed, course or the position.

Data modification can be targeted on the identity or on the coordinates of the vessel. Data removal can affect targeted data fields or whole messages, the frequency of the removed data fields or messages being a variable of the study. In this case, mechanisms for the handling of modified or removed data must be set. In this respect, three main missing data mechanisms are distinguished: missing at random (MAR), missing completely at random (MCAR) and missing not at random (MNAR) [4].

3 An application to maritime scenarios

As defined in [5], the applicative domain for our study is the information that can be extracted from the study of maritime traffic, for which we have parsed AIS messages from an antenna located in Brest, France, as well as antenna, geographic, meteorological, sea state, vessel and port data, all aligned in both time and space with AIS

data [6]. In this end, a set of MSIs have been set, describing a state of a vessel at a given time or interval. Those MSIs encompass speed, course and vessel location features, as well as the state of the AIS transmission and the navigational status. A set of scenarios along the topics of fishing security, sustainable development protection and maritime security [7] have been developed. In this work, the veracity of data varies following the application of methods presented in Section 2, and the consequences on the scenario assessment are evaluated.

Future work will consist in the development and implementation of a methodology for the assessment of the prediction of MSIs and subsequently for the determination of the scenarios, involving *ad hoc* tools for the controlled degradation of datasets (and therefore the measurement of the veracity of a dataset) and the comparison of MSI results that will be the consequence of such dataset variation.

References

1. McAfee, A. and E. Brynjolfsson (2012). *Big data: the management revolution*. In: Harvard Business Review 90(10), pp. 60-66.
2. Iphar, C. (2017). *Formalisation of a data analysis environment based on anomaly detection for risk assessment – Application to Maritime Domain Awareness*. PhD Thesis, MINES ParisTech, PSL Research University
3. Camossi E., A.-L. Jusselme, C. Ray, M. Hadzagic, R. Dréo and C. Claramunt (2017) H2020 EU datAcron project WP-5.3: *Maritime Experiments Specification*. 67p.
4. Schafer, J. and J. Graham (2002). *Missing Data: Our View of the State of the Art*. In: Psychological Methods 7(2), pp. 147-177.
5. Jusselme A.-L., C. Ray, E. Camossi, M. Hadzagic, C. Claramunt, K. Bryan, E. Reardon and M. Ilteris (2016) H2020 EU datAcron project WP-5.1: *Maritime Use Case Description*. 39p.
6. Ray, C., R. Dréo, E. Camossi and A.-L. Jusselme (2018) *Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance* 10.5281/zenodo.1167595
7. Ray C., E. Camossi, A.-L. Jusselme, M. Hadzagic, C. Claramunt and E. Batty (2016) H2020 EU datAcron project WP-5.2: *Maritime Data preparation and curation*. 37p.