Detection of Damaged Fields in AIS Messages ⋆

Marta Szarmach[0000−0002−3793−6641]
and Ireneusz Czarnowski[0000−0003−0867−3114]

Gdynia Maritime University
81-225 Gdynia, ul. Morska 81-87, Poland
m.szarmach@we.umg.edu.pl, i.czarnowski@umg.edu.pl

Abstract. Thanks to the Automatic Identification System (AIS), a telecommunication system designed for maritime purposes, ships from a given area are able to transmit information about themselves (their movement) to other ships, so that neighboring vessels are aware of each others' trajectory and collisions between ships can be avoided. However, due to the technological limitations of AIS (such as the packet collision effect) some of the AIS data are either lost or damaged during the transmission. That is why reconstructing of AIS data is an important issue. In this paper, we propose our method for AIS data reconstruction based on machine learning techniques. Here we mainly focus on the anomaly detection phase, which leads to the identification of incorrect AIS data that has to be reconstructed. Our approach uses clustering to find abnormal datapoints and multi-label classification to distinguish which fields of AIS messages need to be corrected. The preliminary results of such anomaly detection are presented.

Keywords: AIS data reconstruction · Anomaly detection · Multi-label classification

1 Introduction

AIS (Automatic Identification System) is a telecommunication system designed for maritime purposes. It allows ships from a given area to transmit information about themselves and their trajectory to other ships, so that neighboring vessels can be aware of each others' trajectory to avoid collisions between them [5].

AIS consists of two segments. The original segment (terrestrial), utilizing very high frequency radio waves, allows only for communication in a field-of-view (up to 74 km) — ship to ship or ship to shore. To increase the communication

⋆ The project is financed within the framework of the program of the Minister of Science and Higher Education under the name "Regional Excellence Initiative" in the years 2019 - 2022; project number 006/RID/2018/19; the amount of financing PLN 11 870 000 and partially by grant of Department of Information Systems, Gdynia Maritime University.
Our special thanks to Mr Marcin Waraksa and Prof. Jakub Montewka from Gdynia Maritime University for sharing the raw data that we used in our experiment.

range, a new, satellite segment — SAT-AIS — was introduced [3]. In SAT-AIS, a dedicated satellite mediates the communication between many terrestrial areas. However, SAT-AIS struggles againts synchonization issues, mainly packet collision effect [7]: a single satellite covers many terrestrial cells, in which the transmission is synchronised using Self-Organised Time Division Multiple Access, but the cells are not synchronised between themselves, therefore a satellite cannot properly receive data from two or more cells at the same time.

In this paper we propose a machine learning based method for reconstruction of lost or damaged AIS data. One step of the reconstruction is described, namely the anomaly detection phase, which allows identification of AIS data that requires further correction.

2 Proposed Approach of AIS Data Reconstruction

2.1 General Algorithm

The proposed approach of AIS data reconstruction can be divided into 3 steps:

1. **Step one: Clustering.** The AIS data recorded from a given area in a given time is divided into groups, such that each group consists of messages from one and only one vessel.
2. **Step two: Anomaly detection.** The aim of this phase is to identify incorrect data that requires correction: either the whole messages or their damaged parts (fields). This is done in two ways simultaneously:
   - by analysing each group obtained during the clustering phase to find outlying messages,
   - by searching for messages that form a separate group (since they differ from others so much, they might be damaged).
3. **Step three: Prediction.** The correct values of an AIS message fields that are considered damaged are predicted, based on other values from a given group.

2.2 Anomaly Detection Phase

Analysing Standalone Clusters. All messages that form separate clusters are considered outliers. Then the groups that they should belong to (consisting of messages that originated from the same ship) are found. The k Nearest Neighbours algorithm [1] is utilized here: the k-NN classifier is trained on the entire dataset (with the labels being the indices of groups that each message is assigned to) and then in predicts the right index.

Searching for Anomalies Within Clusters. In some situations, the message might not be damaged strong enough to form a standalone cluster. Hence, there is a need to analyse each group to find outlying points inside it. In our framework, the 1-D convolutional neural network is going to be used here, with its input being the waveforms of each field building a single trajectory.

Identyfing Damaged Message Fields. When a damaged AIS message is detected, the exact localization of the disturbance must be established. Since the waveform of values of a given field with a incorrect value in it somehow resembles a wavelet, a wavelet transform [2] of the preprocessed waveform is computed and $\Delta\phi$, the relative difference of its maximum value (for waveforms with and without the potentially incorrect values) is calculated. Also $\Delta\sigma$, the same difference of computed standard deviation for each waveform is calculated. The higher those differences, the more likely the field they were computed on contains an incorrect value. Then the pre-trained Random Forest [4] classifier is given the 2-element vector $[\Delta\phi, \Delta\sigma]$ and provides the multi-label classification, deciding are the differences high enough to consider the field as damaged. Since there might be disturbances in many fields in the entire message, not only one, the classification in this task must be multi-label.

3 Computational Experiment

3.1 Overview

Aim of The Experiment. The aim of the computational experiment is to verify the effectiveness of the proposed method of detecting damaged AIS messages and their fields. In this paper, we present the results of analysing only messages forming separate clusters.

Data. The data used in this experiment is a real data recorded from AIS. Among all 27 types of messages transmitted in AIS, only 3 were used — types 1, 2 and 3, called position reports, which carry the information regarding ship's trajectory [6]. The features extracted from those messages to build the input dataset are: MMSI (ship's identifier), longitude, latitude, navigational status, special manouevre indicator, speed over ground, course over ground, true heading.

The data is gathered into 3 datasets:

- 805 messages from 22 vessels from the area of Gulf of Gdańsk,
- 19999 messages from 387 vessels from the area of Baltic Sea,
- 19999 messages from 524 vessels from the area of Gibraltar.

Methodology. All 3 datasets were examined — for each, 20 randomly selected messages were artificially damaged (1 or 2 of their bits were swapped), then the anomaly detection algorithm was run to find those damaged messages (content of standalone clusters), decide which ships they were originating from (using k-NN with the $k = 5$) and then search for the damaged fields (using Random Forest classifier, $max\_depth = 5$, $n\_estimators = 15$).

3.2 Results

The mean of gathered results of each message is presented in Tab 1.

Table 1. Results of anomaly detection in AIS data in standalone clusters

-	Metric	1. dataset	2. dataset	3. dataset
1 bit damaged	Assignment	96,00%	59,00%	72,00%
	Recall	100,00%	100,00%	99,00%
	Precision	93,33%	70,83%	82,50%
	F1	95,33%	78,33%	87,33%
	Jaccard	93,33%	70,83%	82,50%
	Hamming	98,86%	93,93%	97,00%
2 bits damaged	Assignment	99,00%	68,00%	72,00%
	Recall	81,00%	79,00%	75,50%
	Precision	98,50%	82,00%	86,33%
	F1	86,53%	75,37%	76,13%
	Jaccard	79,92%	64,08%	65,42%
	Hamming	97,00%	93,36%	94,00%

The results are promising — as can be noticed, the recall (which is considered the most impactful metric here, since it is more important to maximize the detection rate of all damaged fields than to minimize the false negative rate) did not drop below 75.5%, mostly varying between 80%-100%. Naturally, the recall was slightly lower for bigger error rate and dataset size.

4 Conclusions

The proposed method of finding damaged AIS messages and their fields seems to provide promising results. However, the work is still in progress — in the near future, the effectiveness of finding anomalies inside groups should be examined, as well as the impact of batch size on the anomaly detection phase quality.

References

1. Altman, N.S.: An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. The American Statistician **46**(3), 175-185 (1992)
2. Debnath, L.: Wavelet transform and their applications. PINSA-A **64,A** (6), 685-713 (1998)
3. European Space Agency Homepage, https://artes.esa.int/sat-ais/overview. Last accessed 25 Feb 2022
4. Ho, T.K.: The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Anaysis and Machine Intelligence **20**(8), 832-844 (1998)
5. International Maritime Organization Homepage, https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx. Last accessed 25 Feb 2022
6. Recommendation ITU-R M.1371-5, https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I!!PDF-E.pdf. Last accessed 25 Feb 2022
7. Wawrzaszek, R., Waraksa, M., Kalarus, M., Juchnikowski, G., Gorski, T.: Detection and Decoding of AIS Navigation Messages by a Low Earth Orbit Satellite. In: Sasiadek, J. (eds) Aerospace Robotics III. GeoPlanet: Earth and Planetary Sciences, pp. 45-62, Springer, Cham (2019)