Detection of suspicious behavior from ship transponder data

Universiteit Leiden
The Netherlands

Michael Bosch (supervised by Frank Takes & Gerrit-Jan de Bruin)

Leiden Institute of Advanced Computer Science

s1875876@liacs.leidenuniv.nl

Societal relevance

'Urk has a problem': Dutch fishing town caught up in cocaine-smuggling trial

Five people, including three fishermen from Urk, are suspected of attempting to smuggle 261 kilos of cocaine in their boat

▲ Urk's isolation, and close-knit community, may have made it a more appealing target for smugglers. Photograph: Judith Jockel for the Guardian

It may be the Netherlands' most religiously devoted community, where television and dancing are spurned by some as the devil's work. But the wrath of God for indulging in those pursuits is unlikely to be the most pressing concern at the moment for some of the 20,000 residents of Urk, for centuries a major centre for Dutch fishing.

Article about Urk, The Guardian, 31 December 2017

Introduction

The news item raised the question whether this was an one-time event or if this is happening more often?

Since there are many ships on The North Sea, the possibilities for smuggling are certainly there. But does this really happen?

And if so, is it possible to automatically detect smuggling and suspicious behaviour using data mining techniques and therefore stop smuggling in the future?

Research question

In this research, data mining techniques on Automatic Identification System (AIS) data and examine if it is possible to detect suspicious behaviour. This research will contribute to a better understanding of AIS data and the the valueable information it contains. To do this, data mining techniques will be used to find patterns in the data and this patterns will be used to detect outliers. Therefore, the research question is as follows:

To what extent can we detect suspicious behaviour from harbors, using data mining techniques on ship transponder data?

Approach

In order to give an answer to my research question, the following steps will be taken:

1. Do **research** in the relevant field to learn more about the automatic tracking system ships use and what kind of data it sends (and thus the data contains).
2. Perform **simple analyses** to calculate averages, plot distributions etc. in order to see what attributes are interesting for further analyses.
3. Analyze the interesting attributes in **more detail**. This can either be:
   (a) Calculate confidence intervals in order to detect outliers
   (b) Use machine learning to predict several characteristics of a ship using e.g. only the route as input

The data

General information

The data that will be used in this research came from the Ministry of Infrasture and Water Management. The data consists of 448 million entries and was collected in several weeks. Moreover, there are 46 attributes for each record, which will be shown in more detail in the next paragraphs. The total size of the dataset is over 91 GB.

Human Environment and Transport Inspectorate
Ministry of Infrastructure and Water Management

Ministry of Infrasture and Water Management

Type of data

Each data entry includes information about the location, speed, call sign, name and heading of the ship as well as the ships' length, breadth and the estimated arrival time at the destination.

Sample of the dataset

	t_trackid	t_starttime	t_updatetime	t_duration	t_callsign	t_mmsi	t_name	t_latitude	t_longitude	t_orientation	t_rateofturn
0	2017033110542822342	2017-03-31 10:54:28.081	2017-04-01 00:00:00.01	3	2ITA4	235112573	ATLANTIC STAR	54.168641033	5.218501	27	

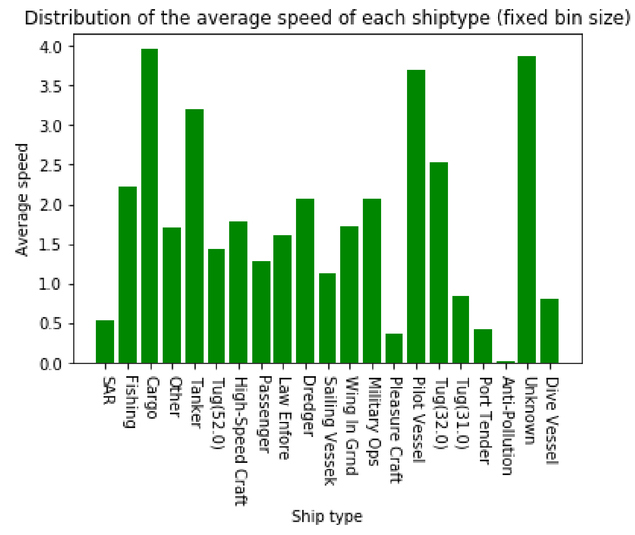
t_length	t_breadth	t_sensors	t_navstatus	t_atonoffpos	t_planid	t_no_orientation	t_org_track_id	t_imo	t_speed	t_heading	t_status_lost
296	38	2	0		2017033114292020731		22342	9670573	9.2	28.2	f

t_status_not_stable	t_status_label_lost	t_freetext	t_planupdatetime	t_vesseltype	p_callsign	p_eta	p_destination	p_name	p_mmsi	p_imo
f	f		2017-03-31 19:15:53.01		2ITA4	2017-04-01 19:00:00	GOTHENBURG	ATLANTIC STAR	235112573	9670573

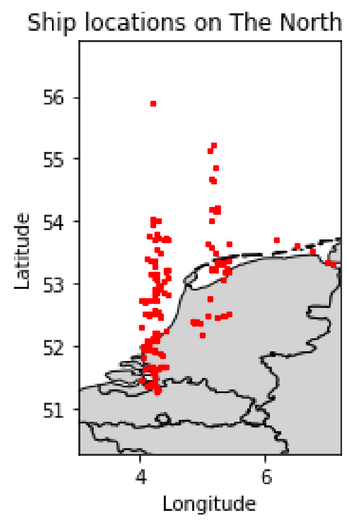
p_length	p_breadth	p_draught	p_sourcename	p_antposfront	p_antposleft	p_shiptype	p_cargotype	p_atontype	p_remark	p_shipid	p_vesseltype
296	38	9.6	AIS	129	35	70	1				

Sample of the dataset

Figures



Distribution of the average speed of each shiptype



Ship locations on The North Sea

Possible outcomes

- A method to detect outliers for a specific numeric value. For example, the duration of a trip based on the route and the type of the ship or the average speed of a ship compared to its type.
- A machine learning model to predict several attributes of a ship based on one or a few attributes. For example, to predict the duration of a trip given only the route.