The automatic identification system of maritime accident risk using rule-based reasoning

Bilal Idiri, Aldo Napoli

HAL Id: hal-00736167

https://minesparis-psl.hal.science/hal-00736167

Submitted on 9 Apr 2013

The automatic identification system of maritime accident risk using rule-based reasoning

IDIRI B.
MINES ParisTech
Centre for Crisis and Risk Research
Sophia Antipolis, France
bilal.idiri@mines-paristech.fr

NAPOLI A.
MINES ParisTech
Centre for Crisis and Risk Research
Sophia Antipolis, France
aldo.napoli@mines-paristech.fr

Abstract - *Current maritime traffic monitoring systems are not sufficiently adapted to the identification of maritime accident risk. It is very difficult for operators responsible for monitoring traffic to identify which vessels are at risk among all the shipping traffic displayed on their screen. They are overwhelmed by huge amount of kinematic ship data to be decoded. To improve this situation, this paper proposes a system for the automatic identification of maritime accident risk. The system consists of two modules. The first automates expert knowledge acquisition through the computerized exploration of historical maritime data, and the second provides a rule-based reasoning mechanism.*

Keywords: Maritime monitoring systems, automatic rule-based reasoning, data mining, maritime accident risk identification.

1   Introduction

Maritime transport is an essential element in economic development and maintaining links with remote territories. Shipping operates in a context that is both spatial-temporal (i.e. the physical environment) and strategic (i.e. deployment modes). The maritime zone is a complex environment that consists of many coastal states with different regulations and involves many different actors. Ships are exposed to the risk of maritime accidents such as sinking or grounding in coastal waters.

With the advent of localisation technologies, telecommunications, technological advances in embedded systems and digital maps, new maritime tracking systems such as the French *Surveillance des espaces sous juridiction nationale et des approches maritimes* (SPATIONAV), the American Integrated Control and Safety System (ICSS) and the Spanish *Sistema Integrado de vigilancia Exterior* (SIVE) have been developed. These

systems make it possible to track vessel movements (position, heading, speed, home port, etc.) in almost real-time. They consist of a data acquisition infrastructure that can capture and transmit shipping data, and an information system for the storage, processing and display of ship movements on a control interface. Despite the widespread use of these systems (for example, all European Union countries have implemented the technology) the number of maritime accidents remains significant. Worldwide, there are still many thousands of maritime accidents each year and in 2008-2009 the amount of oil and hazardous substances accidentally spilled at sea increased by 400% [1]. The most recent accidents include the cruise ship Costa Concordia, which ran aground off the coast of Italy on 13th January, 2012 causing 32 deaths. The vast area to be monitored, the number of ships at sea (41 million ship positions/day for 62,000 ships according to Lloyds) and the multiplicity of scenarios make it very difficult to identify maritime accident risk. Although the timely identification of these risks is crucial, the current context makes this task challenging and complicated.

This paper proposes an approach to the identification of maritime accident risk that is based on automated reasoning, which is in turn founded on knowledge rules. Knowledge rules are obtained through data mining analysis methods applied to maritime databases. The goal is to provide a tool that offers a comprehensive picture of risk events that are scattered and fragmented in time and space, in order to help maritime shipping controllers to understand the latent meaning of the facts displayed on their monitors. Risk events indicate an imminent risk of a maritime accident and the tool can also display alarms that alert operators to events of interest.

Definitions

- Automatic reasoning is a subdomain of Artificial Intelligence[1]. It makes it possible to simulate human reasoning with a computer in order to deduce new knowledge from a sequence of input events and existing knowledge.

- Analogous to the search for nuggets in a goldfield, data mining aims to extract hidden information through a global analysis of large datasets. It seeks to discover models (patterns) that are difficult to identify directly due to the large volume of data, the number of variables to be taken into account and the fact that there are unknown assumptions [2].

2   Background

Several earlier studies have addressed the issue of improving maritime tracking systems – either at the level of the data acquisition infrastructure (the integration of new sensors, sonar networks, aerial drones, etc.) or at the information processing level. Among the work that has been carried out to improve information processing systems, there is the initial Defense Advanced Research Projects Agency (DARPA) project [3], French led DCNS projects [4][5], the work of Vandecasteele and Napoli [6], the work of Etienne, Devogele and Bouju [7] and the research carried out by Roy [8]. The literature also offers several other, more specific studies on the analysis of maritime accident risk. Among these studies, several different approaches can be distinguished: probabilistic [9], statistical [10] and those based on numerical simulation [11]. However, most of these approaches to the modelling and analysis of maritime accidents either do not take into account lessons that can be drawn from historical data or are restricted to confirmatory analyses[2]. This fact makes the discovery, formalization and use of expert knowledge difficult and complicated. Moreover, these approaches make it impossible to discover models that are not already known by experts.

Our research is inspired by the work of Roy [8] who applied rule-based reasoning to the automatic identification of abnormal ship behaviour. However, it should be noted that abnormal ship behaviour does not necessarily help in the identification of maritime accident risk, while normal

---

[1] Artificial Intelligence is a research domain that aims to create intelligent machines. The central problems include the modelling of traits such as reasoning, perception, recognition, etc.

[2] This is an analysis that aims to confirm or negate an initial hypothesis.

behaviour may indicate or pose a risk. For example, the entry or exit of a Ro-Ro (Roll-on/Roll-off) type vessel into or out of a port can be considered normal behaviour, while 76% of accidents involving this type of vessel occur in, or near ports.

3   Methodology

In order to automate the identification of maritime risks we chose a computerised reasoning approach. Our choice was influenced by the fact that this type of approach is much easier to implement than mathematical models. The modularity of the knowledge discovered simplifies system maintenance and makes it easy for users to add or remove knowledge.

Our work is distinct from automation studies of maritime monitoring [3][8] in that it focuses on maritime safety (e.g. grounding, sinking) and not security (e.g. terrorist attacks, illegal immigration).

We chose to use automatic data exploration methods for expert knowledge acquisition. This approach simplifies the task of experts, who only have to validate the knowledge generated. The literature often describes brainstorming methods that bring together domain specialists in order to acquire and define expert knowledge. While these methods are clearly interesting, they are also complicated and expensive to implement and the output scenarios are highly dependent on the personal experiences of the experts. The expert knowledge acquired using our method is not limited to ship behaviour but also includes risk situations and areas. A risk situation is defined as one that meets the conditions (sea conditions, vessel characteristics) conducive to risk, while risk areas are areas where there is a high accident rate.

3.1   Data mining

The extraction of knowledge for the identification of maritime accident risk uses data mining methods that transform historical maritime data into knowledge. This approach solves the problem of knowledge acquisition that is known to be a bottleneck in the process of building expert automated reasoning systems.

In order to maximize the potential for risk identification, the selected approach aims to extract three types of knowledge, namely: risk situations, risk areas and risk behaviours. Risk behaviours in this context describe ship behaviour that indicates a risk (unusual change of course, slowing down, etc.). If conditions do not exist to indicate a risk situation (for example if parameters are missing), the system can draw upon knowledge of risk

behaviours (e.g. entry into a high risk area, abrupt change of course) to identify potential risks.

There are several data mining methods available to extract data related to these different types of knowledge. They include: association rule mining, sequence mining, data clustering and trajectory clustering. Whichever method is used, the knowledge discovered through data mining must be validated by domain experts.

3.2   The data

A consequence of the selected approach to automatic knowledge extraction is that the system implementation must include a data acquisition phase. Broadly speaking, the data acquired can be divided into two categories:

- **Static geo-spatial data**: This data is the result of maritime accident investigations. The data mining analysis of this category of data generates knowledge of risk situations and risk areas. Knowledge related to risk situations links information about navigation conditions (wind, currents), vessel characteristics (type, age, etc.) and maritime risks (type, category, etc.) for the forecasting and targeting of maritime accidents. Knowledge of risk areas can enable closer monitoring of shipping in these areas and planning of intervention and rescue measures.

- **Dynamic geo-spatial data**: This category of data tracks ship movements (position, speed, heading, etc.) and supplementary data describing navigation conditions (wind, current, etc.). The data mining analysis of this category of data discovers knowledge describing the risk behaviour of ships.

3.3   Rule-based reasoning

In an automated reasoning system knowledge is often encoded either as rules (generalized examples) or cases (examples). A rule takes the form 'if *antecedent* then *consequent*' where 'antecedent' and 'consequent' express conjunction or disjunction of instances of database objects. Knowledge provides the link between known information (the antecedent) and the information to be deduced (the consequent) or actions to be executed (e.g. display alarm). A case describes a problem and its associated solution. The case paradigm records details of resolved source cases in a database, which are used to solve new problems (target cases).

Once knowledge has been formulated, several types of reasoning can be applied. Among these, the most

frequently used are deductive reasoning and analogical reasoning. In the case of deductive reasoning, output values are deduced from input values and in the case of analogical reasoning, the new problem is reduced to a problem that has been resolved. The known solution is then adapted to the new problem. Table I shows a simple comparison between case-based reasoning (CBR) and rule-based reasoning (RBR).

Table I: Comparison of RBR and CBR

	CBR	RBR
Knowledge	Case	Case generation
Modularity	Problem	Rule
Problem resolution	Adapted case	Rule application (fast)
Reasoning	Non-deductive	Deductive
Acquisition	Easy (episodic problem solving)	Difficult (how to go about resolving a problem)

The adopted approach used RBR rather than CBR. This decision was based on several factors. First, it is easy to understand as humans often reason in the form of rules (i.e. if condition then action). Second, knowledge rules facilitate modularity as a complex problem can be decomposed into simple rules. Finally, reasoning is deductive and not analogical (analogical reasoning can sometimes lead to erroneous conclusions). The weakness of RBR is linked to the knowledge acquisition problem, which the data mining method described earlier is intended to overcome. Moreover, the data mining approach means that the majority of the rules generated using this method can easily be formalized as knowledge rules.

In a rule-based reasoning system, three essential components are defined: a knowledge base, the facts and an inference engine. To make the analogy with human reasoning, the knowledge base is everything that the human being knows. Facts correspond to their perception of the environment (sight, taste, touch, etc.) and the inference engine is the human's capacity to use logic and make judgments related to knowledge.

As Figure 1 shows, the inference engine continuously monitors incoming events (Facts) to see if there are applicable rules in the rule base (Knowledge). Before executing these rules, the engine must resolve any conflicts (run priority, avoid loops, etc.) that may arise between

them. After execution of the rules selected by the engine, new rules and/or new facts will enrich, respectively, the rule base and the fact base.
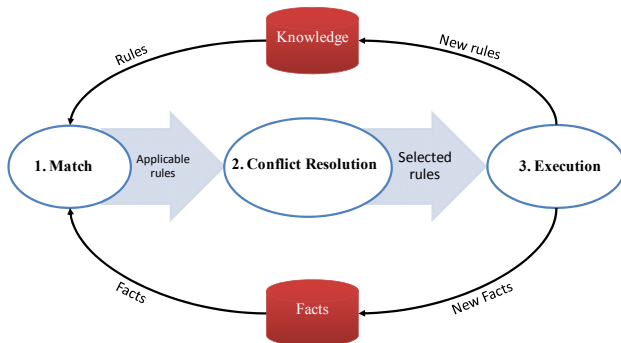

Figure 1. The rule-based reasoning process

4   System Architecture

The system consists of two main modules that are shown in Figure 2. The first module involves expert knowledge acquisition from historical data (accident investigations, ship movements, environment, etc.). The second module puts this knowledge to use in the task of automatic risk identification. The objective is to provide a system that can identify and link, from end-to-end risk events that are scattered and fragmented in time and space, and that can raise alerts to warn operators or experts of potential risks.
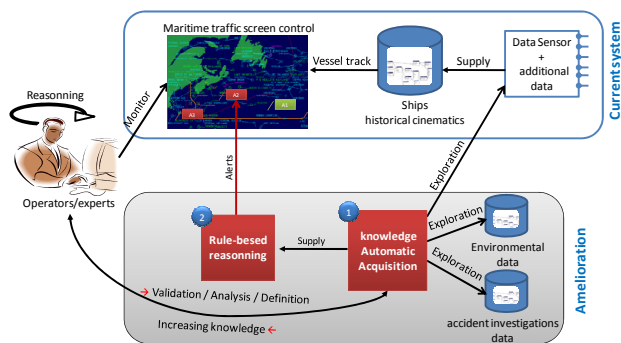

Figure 2. System architecture

The system improves upon existing maritime traffic surveillance systems used by maritime safety organizations (such as the French *Centres Régionaux Opérationnels de Surveillance et de Sauvetage*) in that it helps in risk identification. In addition to the paths taken by shipping that are typically displayed to maritime controllers, the operator is alerted to risk situations and is able to detect shipping that is in risk areas or that displays risk behaviours (slowing down, drifting off course, etc.).

5   Experimentation

In order to test our approach we used a data set from which we were able to extract both knowledge rules and facts. This data was cleaned[3] and pre-processed before being analysed using data mining methods.

5.1   Databases

- **MAIB**: The Marine Accident Investigation Branch (MAIB) database holds details of accidents and incidents that have occurred on ships between 1991 and 2009. This data relates to British vessels located anywhere in the world and all ships in United Kingdom territorial waters at the time of the accident. The database contains records of 14,900 accidents and incidents involving 16,230 ships.

- **MERRA**: Modern-Era Retrospective analysis for Research and Applications (MERRA) is a NASA project that provides meteorological data. Historical data from the period 1991-2009 helped to complete data missing from the MAIB database (winds, surface currents, etc.) and a daily download of weather forecasts made it possible to feed weather data into the fact base.

- **AIS**: The Automatic Identification System (AIS) database provides information about ship kinematics (position, course, speed, etc.) which is transmitted in almost real-time by AIS sensors installed on-board vessels. AIS is mandatory worldwide on all merchant ships over 300 tons and, in the European Union, on all fishing vessels over 15 meters.

5.2   Knowledge acquisition

In an earlier study, we outline how knowledge describing risk situations was acquired [12]. In that study we applied the association rule learning method to the MAIB database. Association rule learning is an unsupervised data mining method that makes it possible to extract knowledge rules of the type 'if *condition* then *result*' from *itemsets* that frequently appear together in a database (an *itemset* is a set of *items* and an *item* is an instance of a database object).

---

[3] Data cleaning involves the identification and correction of anomalies in the data (incompleteness, incoherencies, inaccuracies, etc.).

We applied the Apriori algorithm (developed by Christian Borgelt) implemented in the Rattle 2.6.4 software package (Rattle provides a graphical user interface for data mining) to the pre-processed MAIB data. The result of the analysis was rules that linked known parameters such as the type of vessel, its position, the geographical area and a specific context (e.g. oceanographic and meteorological conditions) with predictive parameters (e.g. type of accident). An example of the rules acquired in this way is shown below (Tableau II).

Tableau II: An example of the rules

Rule	{Location = Coastal waters, Vessel_Category = Fishing/processing, Age_Slice_Of_Vessel = 11 to 18 years} → {Incident_Type = Machinery Failure}
Measures	support[4] = 0.086 confidence[5] = 0.725 lift[6] = 1.47
Interpretation	If there is a fishing vessel, aged 11-18 years, sailing in coastal waters then there is a risk that it will break down.

This rule has been carefully chosen to show the value of combining the various aspects of knowledge (as described in Section 3.1). However, the simple application of this rule in the inference engine will return a huge number of results as many events will meet these conditions. Therefore, before creating an alert, the inference engine must also examine the behaviour of selected vessels (stopping at sea, slowing down, drifting off course, etc.).

5.3 Implementation of rule-based reasoning

Automatic reasoning was implemented using the Drools 5.4.0 (JBoss Rules) software package. Drools is a free, open source and powerful business rule management system. The components of the Drools platform that were implemented were the Business Rules Engine (Drools Expert) and the Business Rules Repository (Drools Guvnor). In our current work rules are added using the

---

[4] The *support* of an itemset is a reliability indicator; it is defined as the proportion of occurrences of the itemset in the database.

[5] *Confidence* is an indicator of the precision of the rule; it is equal to the probability that the consequent will occur together with a particular antecedent.

[6] The *lift* of a rule is an indicator that makes it possible to verify that the results obtained are not due to chance.

Guvnor user interface, however we plan in the near future to integrate all knowledge rules directly through the Java application programming interface (API).

5.4 Example of rule deployment in Drools

Population of the rule and fact base in Drools can be done in several ways: through the Java API, by extending the Drools codebase, or by using the Business Rules Management System (BRMS) interface. The example given below describes the deployment of the rule described in Section 5.2 and shows how to add a fact. This example demonstrates how easy it is to deploy and understand rule-based reasoning.

- **Add rules:**

Rules "Risk of mechanical failure"

when
  $s: ship (location= = "Coastal waters", ship class =="fishing", age ≥ 11, age ≤ 18)

then
  check_behaviour ($s.id_ship);

end

Rules "Behaviour: drifting of course"

when
  $r: risk (Type=="Mechanical failure")
  course (behaviour=="Drifting")

then
  channel["alerts"].send(new Alert());

 end

- **Add a fact:**

Rule "Add Ship Course"
when
then
    insert (new Course ());

end

6 Conclusion

This paper proposes a system for the automatic identification of maritime accident risk based on rule-based reasoning with the aim of improving maritime tracking systems. The system continuously applies acquired knowledge rules to the flow of events describing the movement of ships and changing sea conditions, which makes it possible to identify risks in real-time. The system enables maritime traffic controllers to focus on important events and reduces the complexity of maritime risk monitoring.

The knowledge rules that feed the automated reasoning system are generated using data mining methods. This resolves the problem of knowledge acquisition that is often found to be a bottleneck in automated expert reasoning systems. While brainstorming methods are clearly useful, they are complicated to implement, expensive and the output scenarios rely heavily on the expert knowledge of individuals.

To maximize the potential for risk identification our knowledge discovery method divides knowledge into three aspects that describe: risk situations, risk areas and risk behaviours. The combination of these three aspects improves risk identification.

In the longer term further improvements could be added to the system. These include an alert scheduling system that would prioritise alarms and provide recommendations for intervention measures. We can use Bayesian networks[7] to manage the recommendations for intervention measures as proposed in [13].

References

[1] Centre de documentation de recherche et d'expérimentations sur les pollutions accidentelles des eaux, "Analyse annuelle des déversements de polluants en mer et sur le littoral" 2009. [Online]. Available: http://www.cedre.fr//fr/accident/pollution-accidentelle-mer-littoral.php#combien.

[2] G. Gardarin, *Internet/intranet et bases de données*, Eyrolles edition, pp. 264, 1999.

[3] Darpa, "Predictive Analysis For Naval Deployment Activities (PANDA)", 2005. [Online]. Available: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=7c1de14f8cdfd3d14f112f99a2bbd782&_cview=0.

[4] M. Morel, V. Flori, O. Poirel, A. Napoli, P. Salom, and G. Proutiere Maulion, "Trairtement et Authentification des MenAces et RISques en mer", in Workshop Interdisciplinaire sur la Sécurité Globale (WISG08), 2011.

[5] M. Morel, "SisMaris: Système d'Information et de Surveillance MARitime pour l'Identification des comportements Suspects de navire", in première Conférence Méditerranéenne Côtière et Maritime CM 2, pp. 261-264, 2009.

[6] A. Vandecasteele and A. NAPOLI, "Using Spatial Ontologies for Detecting Abnormal Maritime Behaviour", in Proceedings OCEANS 2012, 2012.

[7] L. Etienne, T. Devogele, and A. Bouju, "Spatio-Temporal Trajectory Analysis of Mobile Objects following", in Proceedings of the International Symposium on Spatial Data Handling (SDH), 2010.

[8] J. Roy, "Rule-based expert system for maritime anomaly detection", in Proceedings of Sensors, and Command, Control, Communication, and Intelligence (C3I) Technologies for Homeland Defense VIX, SPIE, Paper 7666-97, Orlando, F.L, USA, April 2010.

[9] M. D. Amrozowicz, A. Brown, and M. Golay, "A Probabilistic Analysis Of Tanker Groundings", in 7th International Offshore and Polar Engineering Conference, pp. 1-19, 1997.

[10] L. A. LeBlanc and C. T. Rucks, "A multiple discriminant analysis of vessel accidents", *Accident Analysis & Prevention journal*, vol. 28, no. 04, pp. 501-510, 1996.

[11] C. Nishizaki, H. Itoh, K. Yoshimura, K. Hikida, and N. Mitomo, "Development of a method for marine accident analysis with bridge simulator," in 6th International Conference on System of Systems Engineering (SoSE), pp. 31-36, 2011.

[12] B. IDIRI and A. NAPOLI, "Découverte de règles d'association pour l'aide à la prévision des accidents maritimes," in 12ème Conférence Internationale Francophone sur l'Extraction et la Gestion de Connaissance (EGC 2012), Revue des Nouvelles Technologies de l'Information RNTI-E-23 Éditions Hermann, pp. 243-248, 2012.

[13] X. Chaze, A. Bouejla, A. Napoli, F. Guarnieri, T. Eude, and B. Alhadef, "The Contribution of Bayesian Networks to Manage Risks of Maritime Piracy against Oil Offshore Fields," in International Workshops Information Technologies for the Maritime Sector (ITEMS 2012), Database Systems for Advanced Applications, Lecture Notes in Computer Science Volume 7240, pp. 81-91, 2012.

---

[7] A Bayesian network is a system for the representation of knowledge and the calculation of conditional probabilities.