Textual Risk Mining for Maritime Situational Awareness

Amir H. Razavi[1], Diana Inkpen[1], Rafael Falcon[2] and Rami Abielmona[2]

[1]*Electrical Engineering & Computer Science, University of Ottawa*
[2]*Research & Engineering, Larus Technologies Corporation*

***Abstract*— In this paper, we propose an auxiliary Machine Learning (ML) and Natural Language Processing (NLP) integrated system for maritime situational awareness (MSA) operations. We bring into account a new and influential asset – human intuition and perception – to the existing semi-automated decision support systems that mostly rely on numerical data collected by electronic sensors or cameras located either directly on the vessels or in the maritime command-and-control centers.**

For our project, we gathered weekly textual reports spanning twelve months from the United States Worldwide Threats to Shipping Reports repository that belongs to the National Geospatial-Intelligence Agency (NGA). We considered the maritime incident reports written by human operators as a valuable and accessible unstructured textual input source in which a span of text[1] is called "*risk*" if it expresses one of the following kinds of vessel incidents: fired, robbed, boarded, hijacked, attacked, chased, approached, kidnapped, boarding attempted, suspiciously approached or clashed with.

Our approach benefits from probability distributions of some useful features annotated based on a list of lexicons that contain expressions denoting vessel types, risks types, risk associates, maritime geographical locations, dates and times. These distributions are captured and used to anchor the span of "risks" as they are described in the textual reports. After some pre-processing steps that include tokenization, named entity extraction and part-of-speech tagging, the textual risk mining system applies a variety of sequence classification algorithms, e.g., Conditional Random Fields, Conditional Markov Models and Hidden Markov Models in order to compare the risk classification performance. Empirical results show that our NLP/ML-based system can extract variable-length risk spans from the textual reports with about 90% correctness.

***Index Terms*— natural language processing, machine learning, maritime domain awareness, maritime situational awareness, risk detection, text analysis, sequence-based classifiers**

I. Introduction

Information extraction includes research in text mining and Web mining. Its main goal is to extract structured information from unstructured or semi-structured textual input data and has a wide range of applications in a variety of domains, such as business intelligence and biomedical literature mining.

While reviewing the projects and related articles in maritime situational awareness, we observed that a rapid and proper response to a maritime incident coupled with risk management has always been of interest to many governments and organizations. Nevertheless, the lack of one of the most possible influential assets, which is human intuition and perception, can be evidenced in most of them. Almost all of the selected approaches (that will be listed in the next section) only benefit from various numerical data collected by electronic sensors or cameras integrated with structured data gathered from machinery equipment and devices in the vessels or maritime control centers. In this project, we decided to bring the systematic human perception analysis into account and to consider maritime incident reports written by human operators as a valuable and accessible input source to improve situational awareness and maritime risk management in general. Computer software does not have the ability to capture the exact concept of a risk context; yet, there are probability distributions of some useful features that can be captured and used to anchor the span of risks as they are described in the texts of these reports. As examples of these features we can list: vessel type, risk type, risk associates, a maritime general location, a maritime absolute location (e.g., latitude/longitude), date and time. In order to train a classifier and consequently create a model to detect spans of risk in a vast database of textual reports, two human experts (two of the authors) manually annotated the phrases that correspond to risk descriptions in a limited representative subset of reports (52 weekly reports). We also used additional lists of risk features (e.g., risk type, vessel type, location, etc.). Then we automatically annotated some other useful features such as latitude/longitude and named entities in the texts, using regular expressions and named entity recognition (NER) tools. Then, we trained a variety of sequence classifiers on the annotated data that are able to detect occurrence patterns of "risk" and/or "risk factors" in any reports of the same type as the training data.

In order to develop algorithms that extract information from text, we benefit from an NLP package called MinorThird[2] that provides the following capabilities: (1) support for different versions of sequence-based classification algorithms such as: CRF, CMM, HMM; (2) open-source; (3) available for both commercial and research purposes and (4) combination of

[1] Spans are series of adjacent tokens/words that can range in size from a single token/word to an entire document.

[2] http://sourceforge.net/apps/trac/minorthird/wiki/ - Accessed 2013-Nov-15

tools for annotating and visualizing text with a wide range of learning methods.

To the best of our knowledge, there is no existing decision support system (DSS) that extracts risk factors and other related features automatically from textual reports; however there are useful text mining and concept learning approaches for information extraction from other types of texts (that will be discussed and referenced in the next section) including extensions of Conditional Random Fields (CRF), Conditional Markov Models (CMM) and Hidden Markov Models (HMM) that can potentially be combined with other automatic modules for maritime risk assessment.

The rest of the manuscript is structured as follows. Section II reviews some relevant studies while Section III describes the data sets used in this work. Section IV dissects the methodology behind the textual risk mining system and Section V discusses the experimental results. Finally, some concluding remarks are given in Section VI.

II. Related Literature

This section briefly reviews some relevant works in the areas of maritime risk analysis and the mining of textual resources from an NLP standpoint.

A. Risk Analysis in the Maritime World

Maritime Situational Awareness (MSA) is defined as "*the comprehensive fusion of data from every agency and by every nation to improve knowledge of the maritime domain*" by the U.S. National Concept of Operations for Maritime Domain Awareness, in December 2007 [1]. Maritime domain awareness is defined in turn as: "*having true and timely information about everything on, under, related to, adjacent to, or bordering a sea, ocean or other navigable waterway*". This includes all related activities, infrastructure, people, cargo, vessels or other means of transportation. The ultimate goal of marine security is total awareness of anything in the marine domain that could threaten national security[3].

Along with the technologies and equipment that contribute to increased situational awareness in the marine environment, risk detection, risk analysis and risk management stand as pivotal building blocks of any risk-aware DSS. Some successfully deployed examples of DSSs that incorporate risk in maritime situational assessment are: Raytheon's ATHENA [2] (an Integrated Defense System designed to search for suspicious behavior in the maritime search-and- rescue division); DHS' Automated Scene Understanding, a project aiming at interpreting complex information generated by video cameras and other sources in ports across the US [3]; Joint Capability Technology Demonstration (JCTD) [4], a military utility assessment supported by the Joint Requirements Oversight Council and the US Congress; Comprehensive Maritime Awareness (CMA) [5], whose major goal is to share maritime shipping information to prevent similar treats for the commercial maritime shipping; Maritime Automated Super Track Enhanced Reporting (MASTER), an integrative reporting project based on the JCTDs and the CMA and finally, Predictive Analysis for Naval Deployment Activities (PANDA) [6], a case-based reasoning system that uses ontologies for context modeling and business rule representation and evaluation. This system is also equipped with a manual contextual-based risk assessment module which relies on a small risk ontology established by human experts.

The aforementioned modern and complex systems are just a few representatives of the plethora of research efforts recently undertaken in computer-assisted or semi-automated maritime risk analysis. Other relevant contributions include: (1) the design-and-operation selection and optimization framework for maritime risk management put forth by Wang et al. [7]; (2) the Bayesian-based uncertainty handling schemes for maritime risk assessment proposed by Merrick and his collaborators [8] [9] [10] [11] [12] [13]; and (3) the comprehensive integrated approach for the strategic local maritime hubs by Lim and Jau that aims to develop a regional maritime information sharing network that reuses the scattered maritime experimental information across the world [5].

More recently, Jakob et al. [14] reported the results of a project to investigate agent-based techniques for modeling and reasoning about illegitimate maritime activities. The risk posed by these illegal maritime operations is the underlying theme in this research. In [15], the US National Research Council elaborates on rapid reaction technologies (i.e. those that can be matured in six to eighteen months) covering counterterrorism issues. Malik et al. presented in [16] the outcome of their collaborative project with the US Coast Guard that focused on visual analytics of historic response operations and the assessment of potential risks in the maritime environment associated with the hypothetical allocation of Coast Guard resources. Their system identifies high-risk regions through image processing techniques. Falcon and Abielmona introduced in [17] a complementary version of their previous risk management framework by adding automated monitoring and response selection modules. They applied an evolutionary multi-objective optimization algorithm to evaluate each potential search-and-rescue (SAR) response according to a number of conflictive objectives such as cost, latency and casualty probability and hence identify the most promising ones in order to provide timely decision support for a SAR operator.

B. Mining Textual Resources: NLP Approaches

While the examples above show the prominence and momentum that numerical risk analysis has been gaining in the maritime situational awareness realm, when we turn our attention to textual resources analysis as an added value, we do not observe the same mature outcomes. On the one hand, rule-based information extraction (IE) systems like [18] [19] apply manually-generated linguistic extraction patterns to match text and locate information units. Although these patterns perform well on restricted specific domains, it is very labour-intensive to design the extraction rules. On the other hand, since IE inherently includes identifying segments of text that play certain roles, some statistical-based sequence labeling methods such as Maximum Entropy Markov Models (MEMM) [20] and Conditional Random Fields (CRF) [21] can be applied to cope with the problem.

[3] http://www.tc.gc.ca/eng/marinesecurity/initiatives-235.htm - Accessed 2013-Nov-15

Two recent and valuable surveys on IE are those in [22] [23]. They highlight Named Entity Recognition (NER) and Relation Extraction (RE) as two essential IE components. The former aims at finding names of entities such as people, organizations and locations or specific scientific names such as protein or gene names whereas the latter extracts the semantic relationships among different parts of a textual segment (e.g., sentences, paragraphs, or named entities). The best NER and RE approaches rely on statistical machine learning methods [24]. Some examples of such systems are TextRunner [25], Woe [26] and ReVerb [27].

If we break down the IE systems into components such as NER and RE, most of the IE problems can be transformed into classification tasks, which can be approached through standard supervised learning algorithms [25] [28] [29].

Weakly supervised learning methods have recently emerged as an appealing alternative to classical supervised learning schemes in that they can learn with a much smaller amount of training data. As an example, [30] presents a weakly supervised RE method based on a learning paradigm called "distant supervision" that was applied to a large number of known relation instances from very large knowledge bases in order to create the required training data.

Very few risk assessment systems use IE from text. RARGen [31] is one such system: a text-mining-based software that addresses the risk assessment problem by automatically creating and maintaining risk repositories with the goal of extracting a Risk Association Rules (RARs) table from a corpus of risk analysis documents. The risk repositories were built by human experts based on a manually extracted Risk Terms File that contained all distinct words representing risks.

After the analysis of the above literature, we decided to focus on automatic (rather than manual) textual risk mining part which involves the application of sequence-based weakly supervised learning techniques. By weakly supervised or semi-supervised learning, we mean that we applied lexicons and software tools (including NER tools) that partly annotate the data (seen or unseen) automatically; these lexicons and NER tools in turn can be extended and/or updated through a bootstrapping system. In the next step, our method detects the target concept patterns (risk description) and projects them back to the unseen reports. In other words, by applying the created lexicons and the manual risk annotations, we model the span of risk descriptions over a limited number of reports (i.e., the training data); then the model can be applied over unseen data to automatically detect the risk spans. This approach was not employed in the reviewed literature, but we will show that it is a promising approach for the risk span exploration from textual resources such as maritime incident reports. Our goal is to benefit from IE techniques to extract a range of maritime vessel risks from textual reports.

III. TEXTUAL DATA

For this project, we used publicly available data in order to conduct research and analysis on maritime risks. Hence, we decided to focus on the United States' Worldwide Threats to Shipping (WWTTS) weekly reports[4] compiled by the National Geospatial-Intelligence Agency (NGA) in order to carry out our research endeavour. This valuable source issues weekly reports concerning important maritime incidents around the globe. The plain-text reports are published weekly by the Office of Naval Intelligence (ONI), including a summary and details of recent piracy acts and other hostile actions against commercial shipping worldwide, organized by geographic region. The reports also contain any recent developments in the efforts to prevent piracy and prosecute the aggressors. In our study, we selected reports generated in 2012 (52 weeks) for maritime risk extraction and analysis.

IV. TEXTUAL RISK MINING METHODOLOGY

In this study, we consider a span of text as a risk if the main intention is one of the following kinds of reported (in the NGA reports) vessel incidents: fired, robbed, boarded, hijacked, attacked, chased, approached, kidnapped, boarding attempted, suspiciously approached and clashed with.

As a first step toward our goal, we built a few term files (lexicons), including a "maritime risk associates terminology". Constructing the lexicons was a time-consuming, manual search process over related corpora and articles. However, the Risk Assessment Lexicon of the United States' DHS[5] and its Canadian counterpart [32] were used as reliable assets.

The rest of this section elaborates on the different building blocks of the textual risk mining methodology used in this project.

A. *Maritime Risk Components Lexicon*

In order to collect some sort of signals or cues for the risk detection task, we focused on the risks described in the NGA-WWTTS reports. We noticed that for each recorded incident, some risk factors were specified such as: the vessel at risk, the type of risk, the location where the incident occurred, the risk cause/motivation and some risk indicators (e.g., alarm, bomb, RPG-7, hurricane). For each of these risk factor categories, we built a lexicon that potentially plays the role of risk anchor point to explore the span of risks. The NGA-WWTTS dataset became the primary source that we manually exploited for creating the lexicons. However, we also relied upon additional resources and similar datasets to augment the lexicons. The UK's Marine Accident Investigation Branch (MAIB)[6] dataset, National Search and Rescue Manual (including abbreviations and terminologies)[7], Risk Management Fundamentals[8] and

[4] Data can be downloaded at: (Accessed 2013-Nov-15) http://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_64
[5] http://www.dhs.gov/xlibrary/assets/dhs-risk-lexicon-2010.pdf
[6] Data can be downloaded at: (Accessed 2013-Nov-15) http://www.maib.gov.uk/publications/safety_digests.cfm
[7] File can be downloaded at: (Accessed 2013-Nov-15) http://loki.cgc.gc.ca/cansarp/sarmanuals/nsm.pdf

DHS Risk Lexicon 2010[9] are examples of these extra resources.

Using the aforementioned repositories, we manually created the following lexicons: risk types (49 entries marked as <*risk_threat*>; e.g. kidnapped), vessel types (85 entries marked as <vessel>; e.g., tanker), risk indicators (212 entries marked as <*risk_indicators*>; e.g., alarm), risk specific locations (511 entries marked as <*location_specific*>; e.g., southeast of Cotonou) and risk general locations (72 entries marked as <*location_general*>; e.g., Nigeria).

B. *Annotation*

In the next step, we first manually annotated the span of risk descriptions (marked as <*risk*>); then, using the previous lexicons that were created manually (i.e., by human experts), we coded in Java the automated annotation[10] of the corresponding risk factors. Fig. 1 illustrates an incident report with embedded (annotated) labels:

```
<risk>
    <location_specific> BENIN: </location_specific>
    <vessel>chemical tanker</vessel>
    <risk_threat> fired upon, boarded and robbed </risk_threat>
    on october 2nd at 2337 LT while drifting in position
    <position>04:06n;002:51e</position> approximately 136 nm
    <location_specific> southeast of cotonou </location_specific> benin.
    <risk_indicators>pirates</risk_indicators>
    <risk_indicators>armed</risk_indicators>with
    <risk_indicators>automatic weapons</risk_indicators>
    <risk_threat>approached </risk_threat> in two small boats and
    <risk_indicators>boarded</risk_indicators>  the
    <vessel>vessel </vessel> the crew
    <risk_indicators> retreated</risk_indicators>  into the
    <risk_indicators>citadel </risk_indicators>  and stayed there the whole night,
    when they emerged the next day they found that the
    <risk_indicators> pirates</risk_indicators>  had
    <risk_indicators>stolen </risk_indicators>
    <vessel>ship </vessel>  cash
</risk>
```

Figure 1. Example of a textual incident report annotated with risk labels

We automatically added other annotations such as part-of-speech (POS) tags, geographical positions (latitude and longitude), as well as date and time.

In the last stage, one of the authors manually annotated the spans that describe maritime risks in our twelve months of NGA reports. For validation and subjectivity measurements, 20% of the reports were randomly chosen for annotation by another human judge (another author). According to the annotation done by the second annotator, the recall of the initial annotations was evaluated to 95.9%, while their precision was evaluated to 98%. We also computed the Kappa value that compensates for agreement by chance [33]. The Kappa value was 0.699, which indicates a good inter-annotator agreement; therefore, we can consider the annotation to be reliable. Note that in our case, the probability of agreement between judges by chance is small, since we do not have individual (i.e., separate) risk objects and any variable-size text span can potentially be considered a risk.

C. *Debugging and Label Editing*

Debugging the automated annotation programs, editing the lexicons, applying named entity recognizers (e.g., time and date) and part-of-speech taggers in parallel helped us to improve the quality of the annotations (i.e., improvement via semi-supervised bootstrapping annotation). That quality is considered an essential requirement for the training phase in the next stage.

D. *Extraction Learning*

We ran our automated extraction experiments on the textual NGA-WWTTS data that was annotated in the previous stages. For the learning process, we needed to separate training and testing sets. For this purpose, we either randomly split the annotated data into training and testing subsets (e.g., 75% and 25%) or applied the well-known *n*-fold cross-validation procedure (i.e., the classifier is trained on *n-1* folds of the data and tested on the remaining fold, then this is repeated *n* times for different splits, and the results are averaged over the *n* folds). We will report results for both cases.

The learning process was done using a list of *sequence classifiers*. The main difference between the ordinary classifiers such as decision trees and sequence classifiers such as MEMM is that the former work on individual instances and are trained to classify instances of two or more objects whereas the latter are trained based on a sequence of objects. Hence, sequence classifiers need to find the spans (i.e. portions of the sequence with variable length) to be matched to any of the training annotations. Therefore, the baseline for precision and recall of the sequence classifiers is considered as zero.

In order to potentially improve the quality of *risk span extraction*, we also performed pre-processing steps for tokenization and segmentation. As the outcome of these processes, we could extract additional risk-indicative features, including *Unigram, Bigram and Tri-gram tokens* individually assigned to their probabilities of occurrence *inside, outside, ending and starting* of the *span of risks*. The *length of the risk spans* was the last extra feature that was added to the list. Fig. 2 depicts the flow diagram for the entire risk extraction procedure.

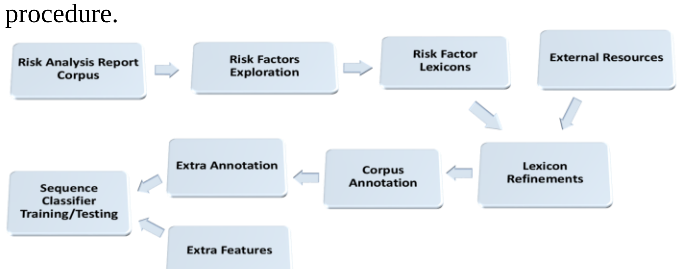

Figure 2. Risk Extraction Architectural Diagram.

V. EXPERIMENTAL RESULTS

We found Precision, Recall and F-measure the most common and declarative evaluation measures recently used in most machine learning papers. The listed performance measures

---

[8] File can be downloaded at: (Accessed 2013-Nov-15) http://www.dhs.gov/xlibrary/assets/rma-risk-management-fundamentals.pdf

[9] File can be downloaded at: (Accessed 2013-Nov-15) http://www.fema.gov/pdf/government/grant/2011/fy11_hsgp_lexicon.pdf

[10] Annotations done in XML format.

were calculated for a variety of sequence classifiers including Hidden Markov Model (HMM) and Conditional Random Fields (CRF) algorithms and their extensions. We ran our experiments with the following methods[11]:

- MEMM: Applies logistic regression/Maximum Entropy to learn a conditional Markov model (CMM) [34]
- SVMCMM: Applies probabilistic SVM to learn a conditional Markov model (CMM) (ibid). This method is analogous to the MEMM learner.
- VPHMM: Applies the Voted Perceptron algorithm to learn the parameters of a Hidden Markov Model (HMM). This method is similar to CRF; however it is often less expensive to train [35]
- VPCMM: Uses the voted perceptron algorithm to learn a "conditional Markov model" (CMM). This method is also analogous to the MEMM learner[12] (ibid)
- CRF (Seq. CRF): implemented based on the IIT CRF [36], in which optimization is performed using the limited-memory BFGS technique of Nocedal and Wright [37].
- CRF using the extra features (Seg. CRF).
- CRf without using the extra features.

We set the number of epochs of all HMMs to 20; also subsequent experiments based on 20, 50, and 100 iterations led us to set the number of iterations for the CRF algorithm equal to 100 in order to achieve the best performance with a running time of less than 30 minutes, over the 5-fold[13] cross-validation on a PC with an Intel i7 processor and 8 GB of memory.

Tables 1 and 2 show the performance of the risk span extraction on two set of experiments run on the NGA data. For evaluation purposes, we chose recall, precision and F1 measures for two different scopes: Token level and Span level. In the first scope (i.e., token), we calculate the listed measures based on the number of tokens that belong to risk spans. In other words, if a token belongs to a risk span and tagged by the classifier as "risk", we count it as a true positive; otherwise, we count it as false positive; the same strategy is taken for the negative side. For the second scope (i.e.. span), we evaluate our method based on whole spans of risks, that means, if our classifiers detect starting, ending and length of a risk span, that will be counted as a true positive; however if even one of the three factors were not exact, we count it as a false positive; again the same strategy is applied for the negative side. The best classifier (e.g., Seg CRF) achieved an F-measure of 0.99 for tokens and 0.90 for exact span matching, based on 75% training and 25% testing set split on the NGA dataset, as the first scenario. The results can be seen in Table 1.

As for the second scenario, we calculated the performance measures on the two scopes using the 5-fold cross validation strategy. This means that we split the entire dataset into 5 equally sized folds with similar class distributions (about 10 weeks of text reports in each fold), then trained a classifier on 4 folds and tested it on the 5th one; this is repeated 5 times on 5 different combinations of train/test folds, and the results are averaged over the 5 runs. Results of the 5-fold cross-validation also confirm the results of our first set of experiments on the split train/test data. They showed that the Seq. CRF classifier achieved an F-measure of 0.98 for tokens and 0.85 for exact span matching. For more details please see Table 2.

TABLE 1. RESULTS BASED ON A RANDOM SPLIT (75% TRAINING AND 25% TESTING SEPARATE SUBSETS)

Evaluation measure → / Sequence Classifiers↓	Token Prec.	Token Recall	Token F1	Span Prec.	Span Recall	Span F1.
MEMM	1.0	0.00521	0.01037	0.0	0.0	0.0
SVMCMM	0.95103	0.97117	0.96099	0.78804	0.78804	0.78804
VPCMM	0.36842	0.17338	0.00123	0.0	0.0	0.0
VPHMM	1.0	0.11732	0.21000	0.33333	0.03571	0.06451
CRF	0.98125	0.99295	0.98706	0.80612	0.80612	0.80612
Seq. CRF	0.98469	0.99543	0.99003	0.81578	0.82010	0.81794
Seg. CRF	0.98687	0.99848	0.99264	0.89312	0.90697	0.90

TABLE 2. RESULTS BASED ON 5-FOLD CROSS-VALIDATION

Evaluation measure → / Sequence Classifiers↓	Token Prec.	Token Recall	Token F1	Span Prec.	Span Recall	Span F1.
MEMM	0.94178	0.12437	0.21973	0.84	0.08805	0.15939
SVMCMM	0.98329	0.97560	0.97943	0.86440	0.85534	0.85985
VPCMM	0.63230	0.90183	0.74339	0.0	0.0	0.0
VPHMM	0.63369	0.88586	0.73885	0.00273	0.01048	0.00433
CRF	0.98291	0.98588	0.98439	0.85894	0.85534	0.85714
Seq. CRF	0.98252	0.98376	0.98314	0.85835	0.85115	0.85473
Seg. CRF	0.98265	0.98034	0.98150	0.84925	0.83857	0.84388

The results showed that our system can detect and extract the span of a variety of maritime incident descriptions from unstructured human textual reports, with acceptable performance. The applied methodology can be potentially applied to detect and extract a wide range of well defined (non-subjective) textual concepts, when the concept is explained with a set of mostly known attributes and components in the content. We can apply the presented approach in other similar domains or tasks (e.g., detecting tsunami or earthquake early reports in local wiki news pages). However, concepts that are more abstract and / or subjective such as "Discrimination" would not work with the proposed approach.

VI. CONCLUSIONS AND FUTURE WORK

A risk mining methodology aimed at drawing meaningful information for maritime surveillance operators from textual reports describing vessel incidents has been proposed in this study. We conducted experiments based on a set of sequence classifiers in order to automatically detect and extract risk spans from the NGA-WWTTS dataset. As a result of our work, the following remarks can be formulated:

[11] The implementations of the above algorithms were drawn from the MinorThird NLP package.

[12] For more details on VPCMM Learner please refer to the javadoc at: http://minorthird.sourceforge.net/javadoc/ (Accessed 2013-Nov-15)

[13] We chose 5 folds to have enough *risk* span entries in each fold.

- The textual risk indicator features can help us automatically extract the maritime risk descriptions out of the human-written reports with acceptable accuracy.
- The lexicons can be considered as the human intuitive core of the risk span extraction learning.
- CRF is the most reliable approach to identify the risk descriptions.
- These risk features could also be automatically extracted and integrated with other numerical data from vessel sensing systems in order to improve maritime situational awareness.
- Human supervision is needed to ensure the generation of high-quality lexicons that consequently affect risk extraction performance.
- The methodology can potentially be applied to similar concept analysis/detection tasks.

In the future, we could use sLDA (supervised Latent Dirichlet Allocation) to automatically extract and augment the risk indicator lexicon. We also plan to experiment with a similar framework upon other maritime risk reporting sources (e.g., the MAIB dataset) in order to compare the quality of risk span extraction across different textual resources and consequently progress toward the projection and augmentation of our methodology onto different corpora in order to extract other types of information.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the National Sciences and Engineering Research Council of Canada (NSERC) through its *Engage* program #439453-12.

REFERENCES

[1] "U.S. National Concept of Operations for Maritime Domain Awareness", December 2007, Available from http://www.ise.gov/sites/default/files/National%20MDA%20CONOPS%202007.pdf – Accessed 2013-Nov-15

[2] N. Friedman "The Naval Institute Guide to World Naval Weapon Systems"; ISBN: 1557502625, 9781557502629; Naval Institute Press, 2006.

[3] "Automated Scene Understanding (ASU)"; Technology and Prototypes Broad Agency Announcement 04-05 (BAA 04-05) Proposer Information Pamphlet (PIP) Department of Homeland Security Homeland Security Advanced Research Projects Agency (HSARPA) April 12, 2004.

[4] Sh. Dewey and J. Maynard; "Time, Technology and Exploitation – can Future Military Command-and-Control Domination be Linked to Effective Knowledge?" Joint Military Operations Department Naval War College, 2008.

[5] I. Lim and F. Jau, "Comprehensive Maritime Domain Awareness: an Idea Whose Time Has Come?" Defence, Terrorism & Security, Globalisation & International Trade Posted: 2007-10-27.

[6] K. E. Moore "Predictive Analysis for Naval Deployment Activities" (PANDA) BAA 05-44, 2005.

[7] J. Wang, H. S. Sii, J. B. Yang, A. Pillay, D. Yu, J. Liu, E. Maistralis and A. Saajedi, "Use of Advances in Technology for Maritime Risk Assessment", Risk Analysis, Vol.24 No.4, 2004.

[8] J. R. W. Merrick, J. R. van Dorp, J. Harrald, T. Mazzuchi, J. Spahn and M. Grabowski, "A Systems Approach to Managing Oil Transportation Risk in Prince William Sound." Systems Engineering 3(3) 128-142, 2000.

[9] J. R. van Dorp, J. R. W. Merrick, J. Harrald, T. Mazzuchi and M. Grabowski, "A Risk Management Procedure for the Washington State Ferries" Risk Analysis 21(1) 127-142, 2001.

[10] J. R. W Merrick, J. R. van Dorp, T. Mazzuchi, J. Harrald, J. Spahn and M. Grabowski, "The Prince William Sound Risk Assessment." Interfaces 32(6) 25-40, 2002.

[11] J. R. W. Merrick, J. R. van Dorp, J. P. Blackford, G. L. Shaw, J. Harrald and T.A. Mazzuchi. "Traffic Density Analysis of Proposed Ferry Service Expansion in San Francisco Bay using a Maritime Simulation Model" Reliability Engineering and System Safety 81(2) 119-132, 2003.

[12] J. R. W. Merrick, J. R. van Dorp and V. Dinesh, "Assessing Uncertainty in Simulation based Maritime Risk Assessment" Risk Analysis 25:3, 731-743, 2005.

[13] J. R. W. Merrick and JR van Dorp "Speaking the Truth in Maritime Risk Assessment" PMID: 16492194 [PubMed], 2006.

[14] M. Jakob, O. Vanek, S. Urban, P. Benda, and M. Pechoucek; "Adversarial Modeling and Reasoning in the Maritime Domain - First Year Report" Agent Technology Center, Department of Cybernetics, FEE Czech Technical University in Prague, 2009, http://agents.felk.cvut.cz

[15] "Experimentation and Rapid Prototyping in Support of Counterterrorism" Committee on Experimentation and Rapid Prototyping in Support of Counterterrorism, National Research Council of National Academies of the United States of America; ISBN: 13:978-0-309-13668-6, 2009.

[16] A. Malik, R. Maciejewski, Y. Jang, S. Oliveros, Y. Yang, B. Maule, M. White and D. S Ebert; "A visual analytics process for maritime response, resource allocation and risk assessment"; DOI: 10.1177/1473871612460991, 2012.

[17] R. Falcon and R. Abielmona, "A Response-Aware Risk Management Framework for Search-and-Rescue Operations" IEEE Congress on Evolutionary Computation, pp. 1540–1547, 2012.

[18] R. Grishman, J. Sterling and C. Macleod. "Description of the PROTEUS System as used for MUC-3" In Proceedings of the 3rd Message Understanding Conference, pp. 183–190, 1991.

[19] W. Lehnert, C. Cardie, D. Fisher, E. Riloff, and R. Williams "Description of the CIRCUS System as used for MUC-3". In Proceedings of the 3rd Message Understanding Conference, pp. 223–233, 1991.

[20] O. Bender, F. Josef Och, and H. Ney "Maximum Entropy Models for Named Entity Recognition" In Proceedings of the 7th Conference on Natural Language Learning, 2003.

[21] B. Settles "Biomedical Named Entity Recognition using Conditional Random Fields and Rich Feature Sets" In Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications, pp. 104–107, 2004.

[22] J. Jiang "Information Extraction from Text" Mining Text Data, DOI 10.1007/978-1-4614-3223-4_2, Springer Science and Business Media, LLC, 2012.

[23] R. Grishman; "Information Extraction: Capabilities and Challenges" Notes prepared for the International Winter School in Language and Speech Technologies - Rovira i Virgili University, 2012.

[24] S. Atdag and V. Labatut "A Comparison of Named Entity Recognition Tools Applied to Biographical Texts". CoRR abs/1308.0661, 2013.

[25] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni "Open Information Extraction from the Web". In Proceedings of the 20th International Joint Conference on Artificial Intelligence, pp. 2670–2676, 2007.

[26] F. Wu and D. S. Weld. “Open Information Extraction using Wikipedia” In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 118–127, 2010.

[27] A. Fader, S. Soderland, and O. Etzioni. “Identifying Relations for Open Information Extraction” In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 1535–1545, 2011.

[28] I. Lewin, I. Silins, A. Korhonen, J. Hogberg, and U. Stenius: “A New Challenge for Text Mining: Cancer Risk Assessment.” In the proceedings of the ISMB BioLINK Special Interest Group on Text Data Mining, 2008.

[29] L. Sun, A. Korhonen, I. Silins and U. Stenius “User-Driven Development of Text Mining Resources for Cancer Risk Assessment” Proceedings of the Workshop on BioNLP, pp. 108–116, Association for Computational Linguistics, 2009.

[30] M. Mintz, Steven Bills, Rion Snow and Daniel Jurafsky. “Distant Supervision for Relation Extraction without Labeled Data” In Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pp. 1003–1011, 2009.

[31] L. Huang, D. Port, L. Wang, T. Xie and T. Menzies, “Text Mining in Supporting Software Systems Risk Assurance” in IEEE ASE’10, pp. 163–166, 2010.

[32] “Intelligence Experts Group All Hazards Risk Assessment Lexicon” Defense R&D Canada, Centre for Security Science; September, 2008.

[33] J. Carletta “Assessing Agreement on Classification Tasks: The Kappa Statistic” Computational Linguistics, 22(2), pp. 249–254. 1996.

[34] A. McCallum, D. Freitag and Fernando Pereira “Maximum Entropy Markov Models for Information Extraction and Segmentation”, Proceedings of the International Conference on Machine Learning (ICML-2000), 2000.

[35] M. Collins “Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms”, Empirical Methods in Natural Language Processing (EMNLP), 2002.

[36] J. Lafferty and A. McCallum and Fernando Pereira “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data”, Proc. 18th International Conf. on Machine Learning, 2001.

[37] F.Sha and F. Pereira “Shallow Parsing with Conditional Random Fields” Proceedings of HLT-NAACL, 2003.