Empirical evaluation of visualizations of normal behavioral models for supporting maritime anomaly detection

Maria Riveiro and Göran Falkman

Abstract—Many approaches for anomaly detection use statistical based methods that build profiles of normality. In these cases, anomalies are defined as deviations from normal models build from representative data. Detection capabilities based solely on these approaches typically generate high false alarm rates due to the difficulty of creating flawless models. In order to support the comprehension, validation, update and use of such models, our latest work has been devoted to the visualization of normal behavioral models of maritime traffic and their usability evaluation. This paper presents the results of a usability assessment carried out in order to evaluate the ability of previously suggested visualizations to support the detection and identification of anomalous vessel behavior.

Index Terms—Anomaly detection, visual analytics, maritime traffic monitoring.

1 Introduction

The surveillance of large sea, air or land areas normally involves the analysis of huge quantities of heterogeneous data from multiple sources, such as radars, cameras, automatic identification systems, etc. In order to support the operator while monitoring large geographical sea areas, the identification of anomalous behavior or situations that might need further investigation may reduce operator's cognitive load.

While it is worth acknowledging that many existing mining applications support identification of anomalous behavior, autonomous anomaly detection systems are rarely used in the maritime domain, due to high false alarm rates, lack of comprehensibility of models used for detection and lack of comprehensibility of system outcomes. Anomaly detection is normally, a complex problem, that can hardly be solved using purely visual analysis nor purely automatic computational methods. Hence, we suggest the adoption of visual analytics [4, 1] principles to support the detection of anomalous behavior in maritime traffic data.

To the best of our knowledge, there has been little previous research assessing the usability of representations of normal models built from data or the usability of visualization of intermediate data mining process steps. This paper presents a usability assessment carried out to evaluate if certain visualizations of normal behavioral models built from maritime traffic data support anomaly detection tasks. We would like to answer the following question: *does the visualization of normal behavioral models built from data support the detection and identification of behavioral anomalies?* This paper builds on previous work presented in [2], where we evaluate if such visualizations support the comprehension of normal vessel behavior (experiment 1). The results of the succeeding second experiment that focuses on the actual detection of anomalous behavior (experiment 2) are presented hereafter.

2 Experiment

In order to carry out representative tasks[1], a proof of concept visual environment, VISAD, has been used. VISAD integrates Google Earth and Matlab modules through an interactive visual interface platform implemented in C#. The module implemented in Matlab includes a particular anomaly detection method (based on a combination of Self-Organizing Maps and Gaussian Mixture Models) that reports for each observation the probability of being anomalous, given the normal model built from the training data set. The vessel traffic data is shown over the geographical map and an integrated module in Matlab displays representations of normal behavioral models. The normal behavioral models capture the dynamic behavior of the different types of vessels: cargo, tanker, passenger, pilot and fishing boat.

Tasks: The tasks carried out by participants cover themes related to the detection, identification and explanation of anomalous behavior from maritime traffic. We divide the exercises into three blocks (numbered starting from the last block of experiment 1):

- Block 5 requests participants to classify certain vessels regarding their normal or not normal behavior (task group compare). Participants should include explanations regarding why they consider the vessels normal or not.
- Block 6 consists of the following exercise: a group of vessels are suggested as anomalous by the system, yet the performance of the anomaly detection capability is reported to be 70%. Participants are asked to identify which of the highlighted vessels are really anomalous and which are just false alarms (task group distinguish and identify). Participants elaborate on their answers, reporting on the rationale behind their choices and confidence.
- Block 7 (post-tasks questionnaire): participants reactions regarding how helpful were the visualizations of normal models in order to support the detection and identification of anomalous behavior are collected. We gather user remarks regarding which tasks were answered using the visualization of normal models provided.

Participants: Twenty two participants, mainly PhD-students, Post-Doctoral researchers and teachers within Computer Sciences, took part in the tests. Seven of the participants were female and the average age was 31.8 years. Participants were randomly divided into two groups, one of them being the reference group. Both groups have to solve the same tasks, but the reference group have no aid regarding the visualization of normal models. They were provided with the same data and the probability of having each observation was displayed for both groups. The model aided group can assess the representations of the normal models and compare particular probability values with the normal model representations (see figure 1).

Metrics: The proposed assessment methodology includes three criteria: *correctness of response*, *time to complete each task* and *user reactions*. The first two criteria measure the ability of the visual representations of normal models to facilitate the detection of hidden anomalies. Both groups are compared regarding the correctness of their responses and the time spent on completing the tasks. The third metric, user reactions, assesses the appropriateness of the visual representations of normal models provided and the extent to which participants view such representations supportive for their goals and tasks. It also includes the interpretation of and participant's level of understanding of such representations. User reactions are collected through a questionnaire with rating and open-ended questions. In addition,

Maria Riveiro and Göran Falkman are with Informatics Research Centre, University of Skövde, Sweden, E-mail: {maria.riveiro, goran.falkman}@his.se.

[1]We have presented a list of maritime traffic monitoring representative tasks for evaluation in [3], after analyzing data gathered during our field work in various maritime control centers

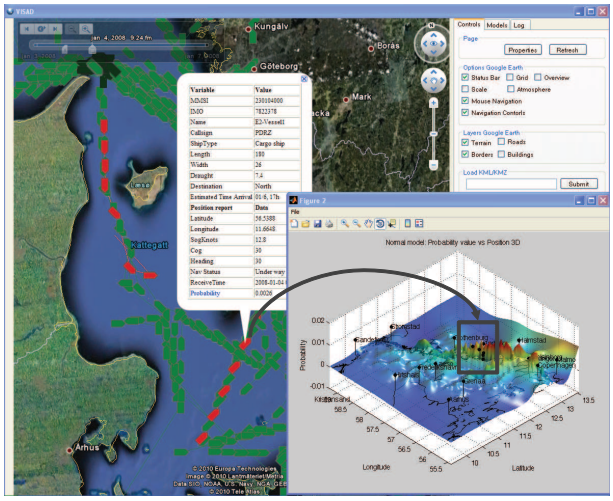

Fig. 1. Visual environment used during the tests. Real AIS data is displayed and analyzed, comparing real-time data to normal models built from the training set. For each observation, a probability value is displayed. When selecting a particular observation, detailed data of the vessel is shown and the normal model window presents the representation of the corresponding model and how the particular probability value (and the five previous observations) relates to the model.

user actions and steps are logged in order to study the strategy that the participants use to solve the tasks.

Data: The data set used for investigative analysis is a subset of real Automatic Identification System (AIS) data, that consist of a collection of real AIS messages broadcasted by vessels traveling along the Swedish west coast, including Gothenburg port area and parts of the coast of Denmark, Germany and Norway. The data corresponds to 17 days during winter. Nine days of the data set are used as training data, i.e. they are considered to model the normal vessel behavior. Eight features are employed for calculating the models, adding the dimensions (length and width) and draught of the vessel to the kinematic features (longitude, latitude, speed, heading and course over ground). Four days of data are analyzed during the exercises by the participants.

3 Results

The analysis of the test data collected during the experiment is organized according to the three metrics considered. In order to compare time and correctness of both groups, a detailed analysis of the data was conducted using a two t-test (significance level $\alpha = 0.05$).

Time to complete each task (block): figure 2 illustrates time spent to carry out the exercises included in block 5 and 6. As shown in the figure, participants aided by visualizations of normal behavioral models spent, in average, more time to solve the tasks at hand. No significant difference was found using t-test $p = 0.7835$ and $p = 0.5025$. Considering the comments written by the participants and the recorded video sessions, this is due to the fact that they spent more time matching the data to normal models and interacting with the visualization of normal models interface.

Correctness of response: figure 3 shows the average number of correct answers per block and group of participants. Participants aided by normal models visualizations performed slightly better than the reference group, a trend that can also be seen in the previous experiment. Nevertheless, p values are not significant, $p = 0.3256$ and $p = 0.3844$. Taking into consideration the time spent to complete each task, even if the group of participants aided by normal models needed more time to carry out the tasks, they performed better, gave more detailed reasons when they categorized vessels as anomalous and were more confident in their decisions.

User reactions: participants rated the helpfulness of the visualizations of the normal models with $3.45/5$ (a Likert scale with five

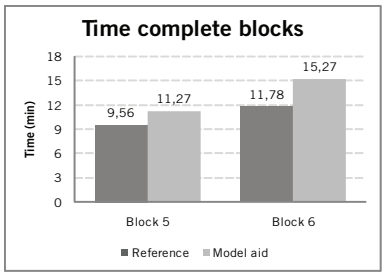

Fig. 2. Average time (in minutes) to complete each of the two blocks that contain exercises related to the detection and identification of anomalous vessels.

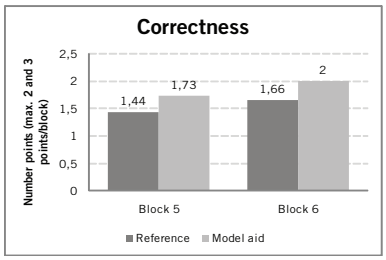

Fig. 3. Correctness of response per block. Block 5 consist of 2 tasks while block 6 has 3 tasks. A task completed with the correct response is given 1, and a task not completed or completed with the wrong response is assigned 0.

ordered response was used, being '5 = strongly agree' with 'the visualization of normal models are helpful to detect and identify anomalous behavior'). Concerning confidence, the qualitative data gathered during the quantitative evaluation (in form of user remarks and explanations) clearly exhibits that participants aided by models are not only more accurate, but also more confident regarding their answers. This is a palpable fact when explanations concerning the abnormality of a particular vessel are given.

4 Conclusions and Future Work

The results of the quantitative evaluation carried out show that participants aided by visualizations of normal behavioral models perform better (taking into account the correctness of response value). No significant difference regarding the time to complete each block of tasks was noted. Current work focuses on a qualitative evaluation that involves the realization of the experiments by three experts in designing and developing military and surveillance systems from Saab Electronic Defence Systems and a group discussion with two experts in maritime surveillance (one from Shipping and Marine Technology, Chalmers University of Technology, Gothenburg and one from VTS West Gothenburg). The qualitative evaluation complements the quantitative approach presented here, since it provides experts insight concerning our suggestions regarding the use of anomaly detection capabilities to support operators in the detection of anomalous behavior and the use of visualizations of normal models.

References

[1] D. Keim, G. Andrienko, J.-D. Fekete, C. Grg, J. Kohlhammer, and G. Melanon. Visual analytics: Definition, process, and challenges. In *Information Visualization*, Lecture Notes in Computer Science, pages 154–175. Springer Berlin Heidelberg, 2008.

[2] M. Riveiro and G. Falkman. Evaluating the usability of visualizations of normal behavioral models for analytical reasoning. In 7^{th} *International Conference Computer Graphics, Imaging and Visualization*. IEEE, 2010.

[3] M. Riveiro and G. Falkman. Supporting the analytical reasoning process in maritime anomaly detection: experimental design. In *14th IEEE International Conference on Information Visualisation (IV10)*. IEEE, 2010.

[4] J. Thomas and K. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, Los Alametos, CA, 2005. [Accessed 2010-10-16].