Chapter 46
The Importance of Visualization and Interaction in the Anomaly Detection Process

Maria Riveiro
Informatics Research Centre, University of Skövde, Skövde, Sweden

ABSTRACT

Large volumes of heterogeneous data from multiple sources need to be analyzed during the surveillance of large sea, air, and land areas. Timely detection and identification of anomalous behavior or any threat activity is an important objective for enabling homeland security. While it is worth acknowledging that many existing mining applications support identification of anomalous behavior, autonomous anomaly detection systems for area surveillance are rarely used in the real world since these capabilities and applications present two critical challenges: they need to provide adequate user support and they need to involve the user in the underlying detection process. Visualization and interaction play a crucial role in providing adequate user support and involving the user in the detection process. Therefore, this chapter elaborates on the role of visualization and interaction in the anomaly detection process, using the surveillance of sea areas as a case study. After providing a brief description of how operators identify conflict traffic situations and anomalies, the anomaly detection problem is characterized from a data mining point of view, suggesting how operators may enhance the process through visualization and interaction.

INTRODUCTION

Exploring, analyzing and making decisions based on vast amounts data are complex tasks that are carried out in a daily basis. People, both in their business and private lives, walk the path from data to decision using diverse means of support. While purely automatic or purely visual analysis methods are used and continued to be developed, the complex nature of many real-world problems makes it indispensable to include humans in the data analysis process.

Automatic analysis methods cannot be applied on ill-defined problems. Furthermore, some real-world problems require dynamic adaptation of the analysis solution, which is very difficult to be handled by automatic means (Keim et al., 2009). Visual analysis methods exploit human

DOI: 10.4018/978-1-4666-9562-7.ch046

creativity, knowledge, intuition and experience to solve problems at hand. While visualization approaches generally give very good results for small data sets, they fail when the required data for solving the problem is too large to be captured by a human analyst (Keim et al., 2009).

The surveillance of large sea areas normally requires the analysis of huge volumes of heterogeneous, multidimensional and dynamic data, in order to improve vessel traffic safety, efficiency and protect the environment (Kharchenko & Vasylyev, 2002). Human operators may be overwhelmed by the data, by the traditional manual methods of data analysis or by other factors, like time pressure, high stress, inconsistencies or the imperfect and uncertain nature of the information. In order to support the operator while monitoring large 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 for area surveillance are rarely used in real world settings. We claim that anomaly detection systems present, among others, two key challenges: they need to provide adequate user support and they need to involve the user in the underlying detection process. Although these aspects cannot be considered independently, they present distinctive characteristics and demand different solutions. The first challenge concerns the necessity of providing adequate user support during the whole detection and identification of anomalous behavior process, allowing a true discourse with the information. This issue includes deepen our understanding of the human analytical and decision making processes. Due to the fact that anomaly detection is a complex and not a well-defined problem, user involvement is needed. The second challenge involves the study of adequate ways of interacting and visualizing the underlying data mining layers. Human expert knowledge is very valuable in these cases, as it can be used to guide the anomaly detection process, for example, reducing the search space, updating knowledge expert rules or refining normal models derived from the data. We believe that the visualization of the data and the data mining process, as well as the availability of interaction techniques play a crucial role in such involvement.

Thus, this chapter aims to: (1) review anomaly detection methods used in the maritime domain, with specific emphasis on the challenges they present from a user's perspective, (2) discuss the role that visualization and interaction plays in the anomaly detection process, (3) identify leverage points where the use of visualization and interaction could make a positive difference, and (4) present examples of how some of the challenges encountered have been tackled in current research carried out at our research center.

The remainder of the chapter is structured as follows: the following section briefly explores the use of visualization and interaction in data mining. The role of visualization and interaction in maritime anomaly detection is discussed afterwards. Then, a review of relevant anomaly detection approaches applied to the maritime anomaly detection problem is presented. Based on field work carried out at various maritime control centers, we provide a brief description of how maritime operators monitor traffic. Enhancements of the anomaly detection process using visualization/interaction and examples are introduced thereafter. Finally, conclusions are outlined.

THE ROLE OF VISUALIZATION AND INTERACTION IN DATA MINING

Data Mining (DM) is defined as the process of identifying or discovering useful and as yet undiscovered knowledge from the real-world data (Hand et al., 2001). Data mining is often placed in the broader context of Knowledge Discovery in Databases (KDD). KDD is an iterative process consisting of data preparation and cleaning, hy-