Chapter 4.5
ASKARI:
A Crime Text Mining Approach

Caroline Chibelushi
Staffordshire University, UK

Bernadette Sharp
Staffordshire University, UK

Hanifa Shah
Staffordshire University, UK

ABSTRACT

The advancement of multimedia and communication systems has not only provided faster and better communication facilities but also facilitated easier means to organized crime. Concern about national security has increased significantly in the recent years due to the increase in organized crimes, leading to increasing amounts of data available for investigation by criminal analysts. The opportunity to analyze this data to determine patterns of criminal behavior, monitor, and predict criminal activities coexists with the threat of information overload. A large amount of information, which is stored in textual and unstructured form, contains a valuable untapped source of data. Data mining and text mining are two key technologies suited to the discovery of underlying patterns in large data sets. This chapter reviews the use of text mining techniques in crime detection projects and describes in detail the text mining approach used in the proposed ASKARI project.

INTRODUCTION

A recent report from the Home Office states that combating organized crime alone costs the United Kingdom (UK) about £40 billion a year (Sandford, 2004). This budget has been used by institutions like the security organizations, law enforcement agencies, and intelligence agencies such as CIA, FBI, and MI5 to dynamically collect and analyze information, and investigate organized crime activities in order to prevent future attacks. These institutions store large amounts of data; recent research has shown that almost 80% of most organizations' information is contained

in text documents (Sullivan, 2001; Tan, 1999), whereas the amount of text/Web mining efforts do not exceed 7% (Drewes, 2002). The speed of security, without information lag, is necessary and requires organizations to make timely and effective decisions. Security organizations acknowledge the need for their textual-based tasks to be organized, managed, and deployed around a set of self-evolving processes, using newly emerging knowledge discovery and agent systems to identify, track, extract, classify, and discover patterns in their corporate databases so that they can be used to generate alerts or crime event notification in real-time. Therefore a clear challenge facing these institutions is how to make effective use of these emerging technologies to assist their intelligence analysts in detecting and anticipating organized crimes, and empower them with powerful tools that can identify patterns, monitor detectable clues across diverse document sources, build behavioral models, and thus improve decision making.

Despite the sudden increase in organized criminal activities in the recent years, there is still no generally accepted definition of organized crime. In order to fight it locally and internationally, we need to understand the common features that characterize the way in which organized criminals operate, as well as how to distinguish organized crimes from other crime. We define organized crime as a (structured or not structured) group of two or more people existing for a period of time and acting in concert with the aim of committing one or more serious crimes that are motivated by politics, religion, race, or financial gain (*Organised Crime in South Africa*, 1998). Organized crime can include terrorism, drug trafficking, fraud, gang robberies, and other group-oriented criminal activities. A terrorist incident is perceived to be significant if it results in loss of life, serious injury to persons, and/or major property damage. Terrorism activities in particular have risen rapidly for the past six years, as shown in Figure 1, which highlights two major incidents between 1998 and 2003. The highest number of casualties is the 1998 attacks in Africa; these attacks included the bombings of USA embassies in East Africa and other different attacks in the region. The second is the September 11, 2001 attacks in the USA. A number of recent attacks have followed namely the bombing of Madrid rail network in May 2004 and the attacks on London transport system in July 2005. These attacks have significantly raised many countries' concerns about national security.

This proliferation of organized crime and the threat of global terrorism have led to the ever-growing volume, variety, and complexity of data captured for analysis. Some intelligence data sources are growing at the rate of four petabytes per month now, and the rate of growth is increasing. The challenge of today lies no longer in the storage and retrieval of data, but in our ability to scan through huge amounts of information and extracting the right information for the right person at the right time.

CRIME PREVENTION AND DETECTION APPROACHES

The concern about national security and crime prevention has increased significantly over the last few years, and has led to the development of national and international funding initiatives, networks and research projects aimed at fighting organized crime. In the USA, the Defence Advanced Research Project Agency (DARPA) has initiated a homeland security program named Total Information Awareness (TIA), which incorporates a number of technologies such as data fusion, database searches, biometrics, and pattern recognition. This program seeks to develop a network of technologies to help security officers predict and prevent terrorism activity (Kenyon, 2003). In the UK, the Engineering and Physical