Chapter 12
Classification of Complaints for Criminal Intelligence Purposes

Pauline Rousseau
CY Cergy Paris University, France

Dimitris Kotzinos
CY Cergy Paris University, France

Daniel Camara
Pole Judiciaire Gendarmerie Nationale, France

ABSTRACT

The increase in the volume of available data is changing how people perceive their own fields and how the people may interact with this surplus of information. Public security is not different; law enforcement agencies (LEAs) now have available a large quantity of information to help them fight criminality. One challenging problem is to classify/predict criminal activities. The differentiation over two different complaints may only be clear through the careful analysis of complaints' open text fields (e.g., the modus operandi), where the specificity of the perpetrated crime is described. Sometimes the intention behind a crime is not evident unless it is correlated to other crimes and patterns get extracted from them. This chapter shows that it is possible to classify criminal data using machine learning-based methods and that open text fields, such as the modus operandi, may play a fundamental role in the performance of the classification.

INTRODUCTION

The increase in available information impacts all aspects of society, and criminality analysis is not an exception. The access to recent technologies and the systematic digitalization of all aspects of the investigative process has largely increased the amount of available information for investigations. This enables a whole new set of possibilities for analyzing and correlating the available data, in ways investigators have never thought of before. Law enforcement agencies (LEAs) have the possibility of performing much more precise and fine-grained analyses of criminality, that were impossible some years ago. One example is

DOI: 10.4018/978-1-6684-5624-8.ch012

the classification of criminal complaints. A complaint is a declaration by a victim over a possible crime. One example of a complaint could be a car's exhaust pipe theft. In this case the car's owner can go to the local police authority and file a complaint to report the car degradation.

Typically, complaints are already classified into broad categories when filed. This classification is used for statistical purposes and for directing the complaints to specialized units for investigation. However, these classifications are quite loose and may not be precise enough to be helpful in criminality understanding, i.e. intelligence purposes. They may also be too broad to be useful for the specialized units, which typically work with more fine-grained classifications inside their broad activity. When trying to understand criminality, agents need a more accurate classification to detect specific phenomena. For example, the objective of the degradation could be not the exhaustion pipe itself, but the precious metal that was within it. In this case units, other than vehicle-related ones, may also be interested and concerned by the complaint. This is even more true if the phenomena are new or not yet widely known from the terrain units, that collect the complaints. Moreover, the complaints' differentiation may only be perceptible over the open text fields, e.g., the description of the crime, called modus operandi (MO). Sometimes specialists can evaluate a crime only through the analysis of the modus operandi. For example, the objective of the vehicle degradation was to collect precious metals, which is something new in that specific region. Even if for a specialist in a specific domain, the distinction between the different sub-classification is relatively simple, the total number of precise sub-classes for all units, and above all, the raw number of entries to classify is potentially huge. Given the volume of data (1.8 million complaints per year), manual classification is not an option.

This chapter evaluates the use of artificial intelligence-based methods to perform a sub-classification of complaints. Complaints consist inherently of semi-structured data, i.e., some fields are structured (e.g., date of the fact, initial classification, the value of the damage/stolen good, age of the victim), others are open free text (e.g., Modus Operandi and the Description of Stollen Goods). The method presented in this work considers that all the available information may be important for a precise binary classification of criminal activities. In general, classification methods either take structured or non-structured data into account, this chapter argues that both are needed. The experimentations will show that using both types of data, a significant performance improvement may be reached. For LEAs the explainability of the method is of paramount importance. This proposal considers explainable and non-explainable methods for comparison.

This chapter is organized as follows; Section I (Background) presents the state of the art in criminal data classification. Section II (Dataset Description) explains the type of data used for classification. Section III (Data pipeline) explains the full data pipeline treatment put in place for performing the classification. Section IV (Classification Methods) explains the methods used for the classification. Section V (Experimentations) presents the results of the classification experiments. Section VI (Conclusion and Future works) presents the conclusions and possible future works on the domain of criminal data classification.

BACKGROUND

A simple way to understand criminality is given by the problem analysis triangle (Clarke, 2005), also called the crime triangle. Accordingly, to this theory, three elements must exist for a crime to occur: an offender, a victim, and a location where the path of the first two crosses. In some sense, the interaction