Available online at www.sciencedirect.com
SCIENCE DIRECT®
Decision Support Systems 41 (2006) 604–615
Decision Support Systems
www.elsevier.com/locate/dsw

An outlier-based data association method for linking criminal incidents

Song Lin*,[1], Donald E. Brown

Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22904, USA

Available online 2 October 2004

Abstract

Serial criminals are a major threat in the modern society. Associating incidents committed by the same offender is of great importance in studying serial criminals. In this paper, we present a new outlier-based approach to resolve this criminal incident association problem. In this approach, criminal incident data are first modeled into a number of cells, and then a measurement function, called outlier score function, is defined over these cells. Incidents in a cell are determined to be associated with each other when the score is significant enough. We applied our approach to a robbery dataset from Richmond, VA. Results show that this method can effectively solve the criminal incident association problem.
© 2004 Published by Elsevier B.V.

Keywords: Outlier detection; Similarity matching; Entropy; Information measures

1. Introduction

Data mining is a collection of techniques that can be used to reveal underlying relationships in a large amount of data. Various data mining approaches have been introduced to the crime analysis field and these have enabled crime analysts to perform some tasks more effectively than ever before.

However, many important problems remain for crime analysts that have not yet been addressed by data mining technologies. Among these is the problem of associating incident reports for crimes perpetrated by the same criminal or criminals. The most compelling examples of this problem involve associating incidents from serial or career criminals. Refs. [5,14] contain a discussion of serial criminals and an example is provided by the recent Washington, DC sniper [30]. For serious part I crimes analysts typically devote as much time as necessary to make the correct associations between incidents. Less serious crimes receive less attention and frequently go without any association at all. The failure to associate these records seriously impedes law enforcement's ability to recognize criminal patterns and make arrests.

Several different methods have been proposed and developed to resolve this criminal incident

* Corresponding author.
E-mail address: sl7h@almuni.virginia.edu (S. Lin).
[1] Home address: 412 Fisher Ave., Piscataway, NJ 08854, United States.

0167-9236/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.dss.2004.06.005

association problem. In agencies where incidents are not well indexed, crime analysts make associations by manually comparing the incident records. For major crimes in the jurisdictions of these agencies, analysts will sometimes spread out the paper copies of the incident reports over several tables and stack the ones that seem to associate. Agencies with usable records management systems use a more automated version of this procedure by retrieving records using a series of Structured Query Language (SQL) queries. Instead of stacks on a table, they store and organize results from their queries in folders.

Other more automated approaches have been developed, but not widely used. The Integrated Criminal Apprehension Program (ICAP) [21] enables police officers to match the suspects with the arrested criminals using Modus Operandi (MO) features. In the Armed Robbery Eidetic Suspect Typing (AREST) program [2], an expert system approach is used. A potential offender can be classified into three categories: probable suspect, possible suspect, and non-suspect. The Violent Criminal Apprehension Program (ViCAP) [22] developed by the Federal Bureau of Investigation (FBI) is an incident matching system. MO factors are primarily considered in ViCAP. In the COPLINK project [10] undertaken by the researchers at the University of Arizona, a concept space model is used to link records in the database with given search terms. Brown and Hagen [7] developed similarity based methods for incident association. In testing with crime analysts, they showed that these methods outperformed the commonly used SQL query in terms of both effectiveness and efficiency.

The Brown and Hagen methods present the most formal approach to data association using similarity measures but each of the above methods either implicitly or explicitly uses measure of similarity or dissimilarity. The theoretical basis for "similarity-based" approaches derives from results in criminology. According to the rational choice theory [11], criminals make decisions that maximize their expected return. Brantingham and Brantingham [5] claim that the environment in which criminals operate contains signals or cues (physical, spatial, cultural, etc.), and criminals use these cues to evaluate their targets and make their decisions. According to this theory, a criminal incident is the outcome of a decision process involving a multi-staged search in the awareness space of the criminal. During the search phase, the criminal associates cues, clusters of cues, or cue sequences with a "good" target. From a crime analyst's standpoint, these cues form a template of the criminal, and once the template is built, it is self-reinforcing and relatively enduring over the time intervals considered. Due to the cognitive limitations of human beings, a criminal normally does not have many decision templates. Therefore, when we observe criminal incidents with the similar temporal, spatial, and modus operandi features, this suggests the incidents come from the same template of the same criminal. Because of these arguments, linking incidents with similar characteristics or templates together appears to be a natural solution to finding serial criminals.

At some level of detail the template used by each criminal is unique to that criminal. Unfortunately, the data collected by the police department typically does not contain enough detail to uniquely identify a template. Given observed (recorded) attributes, some templates are "popular" or "common" and one of these common templates may be shared by different criminals. Hence, criminals with common templates are not differentiable. In these cases, linking incidents solely based upon their similarity may lead to erroneous decisions. On the other hand, some templates are "unusual". For "unusual" or "uncommon" templates, we are more confident in saying that incidents come from the same criminal.

As an example, consider the weapon used in a robbery incident. We may have many incidents with the value "gun" for weapon used. However, no crime analysts would say that the same criminal committed all of these robberies because "gun" is a common template shared by many criminals. If we observe several robbery incidents with an uncommon template, say a "Japanese sword", we are more confident in asserting that these incidents result from the same criminal. (This "Japanese sword" claim was first given in Ref. [7].)

From the above discussions, we see that a good association method should consider not only whether a group of incidents are similar to each other (they result from the same template), but also whether this

Download English Version:

https://daneshyari.com/en/article/553960

Download Persian Version:

https://daneshyari.com/article/553960

Daneshyari.com