Towards the Ordering of Events from Multiple Textual Evidence Sources

Sarabjot Singh Anand, University of Warwick, UK
Arshad Jhumka, University of Warwick, UK
Kimberley Wade, University of Warwick, UK

ABSTRACT

In any criminal investigation, two important problems have to be addressed: (1) integration of multiple data sources to build a concise picture of the events leading up to and/or during the execution of a crime, and (2) determining the order in which these events occurred. This paper focuses on the integration of multiple textual data sources, each providing a recollection of events observed by eyewitnesses. From these textual documents, using text mining and natural language processing techniques the authors identify events (across the document corpus) associated with the crime and infer temporal relationships between these events to create a (partial) ordering of the events. The authors evaluate their method on data collected through a mock eyewitness task.

INTRODUCTION

In any criminal investigation, the re-creation of the order in which events took place leading to a crime is capital. Capturing that order is difficult due to several problems, the main ones being: (1) information about the events is distributed across different sources, and (2) the often heterogeneous nature of the information. For example, information can be video, such as from CCTV cameras, textual information from emails, photographs on social network sites, and structured data from logs and databases. These data have to be merged so that meaningful conclusions can be drawn about the crime.

In this paper, we consider only textual data, coming from different sources. Such a situation arises when individual reports about a crime are gathered, for example, from different eyewitnesses. Individual accounts of events leading to a crime can differ in a number of ways. Firstly, people naturally use different vocabulary. They may also use different tenses when describing events. Secondly, people may be privy to only a subset of the events associated with the crime and, hence, may only provide accounts on the parts that they have witnessed. Thirdly, interested parties may intentionally provide incorrect
information to mislead the investigation. Indeed groups of individuals may collude to confound the investigation. Eyewitnesses may also be misled by investigators. Fourthly, individuals may provide different levels of detail about the observed events. All these facts make the automated extraction of events, and the subsequent matching of events across different textual documents and the identification of actors associated with these events a difficult problem. Thus, a methodology is needed that will, first, discover and then order the events of interest.

**Event Discovery**

To illustrate some of the difficulty in discovering events, we consider some examples obtained from the mock eyewitness task that is used to validate the method presented in this paper: Three simple events that were observed are (1) the robber flicking through a photo album, (2) shutting the door to the garden and (3) taking a coke from the fridge. These common events have been differently described by our subjects, as shown below:

- **Event: Flicking through a photo album**
  - “Watched TV and flicked through a photo album”
  - “Opens the TV and watches some pictures in a photo album”
  - “Switches the TV on and flicks through photo albums whilst sitting on a sofa”

- **Event: Shuts the door to the garden**
  - “He shuts and locks the back doors of the kitchen”
  - “He then closes and locks the door to the garden”
  - “Fixed the light, after this he closed the back door”

- **Event: Taking a coke from the fridge**
  - “Goes to the kitchen where he drinks a Coca-Cola (he takes it from the fridge)”
  - “Helped himself to some coke in the fridge”
  - “Got a can of coke from the fridge before taking an apple and eating it”

It is the objective of the paper to develop an algorithm that would identify these ten descriptions as three distinct real-world events, and to infer temporal ordering relationship between these three events. The above captures the inherent difficulty of determining relevant events from a set of textual data sources due to (1) the disparity in the description of common events, (2) the level of details for each event, and (3) the context each event is set against.

**Event Ordering**

In the context of determining the order in which a set of events occurred, the order of the events, as identified in a given document provided by one eyewitness, may yield a linear structure, i.e., the order in which the events of interest occurred is completely mapped. However, contradictions can arise among various documents, leading to uncertainties, which then yield a branched structure, where each branch represents a possible interleaving of events, i.e., a possible order in which events could have occurred. For example, consider three events $e_1 \cdot e_2 \cdot e_3$. If witnesses $w_1$ and $w_2$ report the order to be $e_1 \cdot e_2 \cdot e_3$, then the structure is a chain. On the other hand, if $w_1$ reports $e_1 \cdot e_2 \cdot e_3$, while $w_2$ reports $e_1 \cdot e_3 \cdot e_2$, then a tree structure is obtained, with node $e_1$ having two branches. Furthermore, not all events within a single document may have temporal ordering relationships that can be inferred from text. Hence, we need to piece together fragments of temporal ordering to obtain one consistent temporal ordering across events.

**Space Complexity**

Moreover, the size of the structure can be very big, depending on the number of events, and documents. For example, if a document $D_i$ contains $e_i$ events, and there are $r$ documents, i.e.,
Cancellable Biometrics for On-line Signature Recognition
www.igi-global.com/chapter/cancellable-biometrics-line-signature-recognition/52860?camid=4v1a

Calm Before the Storm: The Challenges of Cloud Computing in Digital Forensics
www.igi-global.com/article/calm-before-storm/68408?camid=4v1a