A Model Based Reasoning Approach for Generating Plausible Crime Scenarios from Evidence

Jeroen Keppens
Centre for Forensic Statistics and Legal Reasoning
The University of Edinburgh
Appleton Tower, Crichton Street
Edinburgh, EH8 9LE, UK
jeroen@inf.ed.ac.uk

John Zeleznikow
Centre for Forensic Statistics and Legal Reasoning
The University of Edinburgh
Old College, South Bridge
Edinburgh, EH8 9YL, UK
john.zeleznikow@ed.ac.uk

ABSTRACT

Robust decision support systems (DSSs) for crime investigation are difficult to construct because of the almost infinite variation of plausible crime scenarios. Thus existing approaches avoid any explicit reasoning about crime scenarios. They focus on problems such as intelligence analysis and profiling. This paper introduces a novel model based reasoning technique that enables DSSs to automatically construct representations of crime scenarios. It achieves this by storing the component events of the scenarios instead of entire scenarios and by providing an algorithm that can instantiate and compose these component events into useful scenarios. This approach is more adaptable to unanticipated cases than one that represents scenarios explicitly because it allows component events to match the case under investigation in many different ways. The approach presented herein is applied to and illustrated with examples from an application of the differentiation between homicidal, suicidal, accidental and natural death.

Keywords

Model Based Reasoning, Abductive Reasoning, Decision Support Systems, Crime Investigation

1. INTRODUCTION

Two of the most difficult, yet crucial, tasks in crime investigation are hypothesis formulation and evidence collection. Hypothesis formulation is the postulation of plausible properties of the crime under investigation, such as the perpetrator of the crime and his modus operandi. Evidence collection involves the accumulation of observable consequences of the events constituting or surrounding the crime in an effort to prove which of the postulated hypotheses is correct.

These tasks are crucial because failure to consider certain hypotheses or pieces of evidence can cause miscarriages of justice in any subsequent legal proceedings. They are difficult because an investigator's knowledge about what scenarios might, hypothetically speaking, have produced the available evidence is inevitably incomplete. By extension, it is also hard to foresee what evidence collection strategy is most likely to reduce the set of plausible scenarios. Human agents, in particular find it difficult to consider multiple alternative hypotheses simultaneously and tend to try and confirm a single hypothesis instead.

Decision support systems (DSSs) may be able to assist crime investigators in this effort. But, as argued in [15], conventional DSS approaches are not particularly suitable for solving this problem due to their lack of robustness (i.e. the flexibility to deal with unforeseen cases). Each major crime scenario potentially consists of a unique set of circumstances whilst many conventional AI techniques have difficulties in handling previously unseen problem settings. Approaches devised to be adaptable to new situations, such as case based reasoning, tend to work on the assumptions that at least knowledge about settings of a similar specification and with a similar solution to the unseen case are available. This is not the case in major crime investigation. Firstly, certain types of major crimes, e.g. homicides, are extremely rare compared to the occurence of other crimes and other scenarios, e.g. accidental deaths and suicides, that potentially produce similar sets of evidence. Secondly, certain combinations of subtle differences between cases, e.g. the type of relationship between a witness and a suspect, can have a significant impact on a particular case.

This paper aims to be a first step in the development of a DSS to aid crime investigators in hypothesis formulation and evidence collection. A novel model based reasoning technique, derived from the existing technology of compositional modelling [8, 14], to automatically generate crime scenarios from the available evidence will be presented. Consistent with existing work on reasoning about evidence [18, 16], the method presented herein employs abductive reasoning. That is, the scenarios are modelled as the causes of evidence and they are inferred based on the evidence they may have produced.

The approach also uses the notion that unique scenarios consist of more regularly recurring component events that are combined in a unique way. It works by selecting and instantiating generic formal descriptions of such component events, called scenario fragments, from a knowledge base, based on a given set of available evidence, and composing them into plausible scenarios. This approach addresses the robustness issue because it does not require a formal representation of all or a subset of the possible scenarios that the system can encounter. Instead, only a formal representation of the possible component events is required. Because a set of events can be composed in an exponentially large number of combinations to form a scenario, it should be much easier to construct a knowledge base of relevant component events instead of one describing all relevant scenarios.

The remainder of this paper is organised as follows. First, the domain of decision support in crime investigation is surveyed. Then, the decision problem addressed here is formalised. Next, section 4 describes the proposed model based reasoning approach and illustrates it by means of an example. Section 5 discusses the advantages and disadvantages of the proposed approach and shows how it will be used as part of a larger decision support system to be developed in future work. Finally, the paper is concluded in section 6.

2. BACKGROUND

Crime investigation is a complex problem, involving the collection and maintenance of large amounts of data and expert knowledge. As a result, a significant body of research has focused on the development of decision support systems (DSSs) to aid law enforcement agencies in this task. Although a detailed literature review is beyond the scope of this paper, this section presents a brief overview of the domain.

One group of DSSs formalise expert knowledge in the form of a conventional expert system [19]. For example, AREST [1] is an expert system designed for profiling suspects of armed robberies. InvestigAide B&E [22] is an expert systems designed to support the processing and investigation of breaking and entering cases. It supports activities such as gathering and recording case data and provides useful information such as suspect characteristics and similar cases.

Another group of DSSs apply knowledge discovery and data mining techniques to databases containing past cases, police reports and intelligence data. The approaches employed range from data visualisation [10] to the use of more formal statistical analysis [6]. Good examples of mature applications in this area include the COPLINK suite of tools [3, 11] and RECAP [2]. COPLINK is a tool aimed at providing an information extraction facility that integrates data from multiple police forces. RECAP (REgional Crime Analysis Program) is a tool that seeks out patterns of similar modus operandi in an effort to identify organised crimes.

A third group of systems employ Case Based Reasoning (CBR) methods to help investigators discover similar past cases and solution methods that correspond to those past cases. In the context of crime investigation, CBR systems usually perform analysis tasks by means of predefined sets of information. Typical tasks include the categorisation of the risk of electronic commerce transactions [12], the categorisation of crimes and retrieval of cases with similar profiles in burglary [20], and the differentiation between hostile intrusions of computer systems and other anomalous transactions [7].

These systems cover a wide range of problems and employ many different types of inference procedures: including deductive, inductive and abductive ones. However, it can be noted that the vast majority of research into DSSs in the domain of crime investigation existing approaches mainly provide support for analysis of intelligence data and the formalisation of best police procedures. They have avoided the important problem of hypothesis formulation and the overall management of the investigation process. This is understandable (but clearly not desirable) as conventional knowledge based systems, which underlie the existing work, lack the *robustness* needed to cope with the variety of circumstances that may be encountered during investigation.

Generally speaking, systems are said to be robust if they remain operational in circumstances for which they were not designed. In the context of crime investigation systems, robustness requires an adaptability to unforeseen crime scenarios. This objective is difficult to achieve because low volume major crimes tend to be virtually unique.

The problem of finding and reasoning with theories with respect to evidence has been the topic of important existing work within the Artificial Intelligence and Law community. A detailed survey of forms of evidential reasoning is presented in [21]. Recent work, such as [17] and [23], has advanced this field by devising knowledge representations and mechanisms to accurately represent and reason with legal arguments.

The main goal of this work is the construction of an intelligent agent that aids in crime investigation efforts by constructing and analysing a space of alternative theories with only limited user intervention. Therefore, this paper will employ a representational framework that is far less sophisticated than the ones developed in the aforementioned work on evidential reasoning, whilst concentrating on the construction of algorithms that provide the functionality required to meet the challenges that were set out in the introduction.

3. THE DECISION PROBLEM

In crime investigation, it is natural to reason about *crime scenarios*: a description of states and events, changing those states, that may have happened in the real world. The ultimate aim of a crime investigation is to discover the truth or falsehood of certain important properties of the crime scenarios, called *hypotheses* herein. The hypotheses include the possible types of crime (e.g. homicide vs. suicide) or the plausible perpetrators. In order to discover whether certain hypotheses are plausible, crime investigators collect evidence. Pieces of *evidence* are consequences of a crime scenario that are either directly observable, or observable via certain examinations (e.g. forensic tests).

The goal of the DSS described in this paper is to find the set of hypotheses that follow from scenarios that support the entire set of available evidence. This set of hypotheses can be defined as:

$$H_E = \{h \in H \mid \exists S \in \mathbf{S}, (\forall e \in E, (S \vdash e)) \wedge (S \vdash h)\}$$

where H is the set of all hypotheses, $\mathbf{S}$ is the set of all consistent crime scenarios and E is the set of all collected pieces of evidence.

As described in [15], such information can be employed to search for suitable evidence collection strategies. However, a detailed description of this topic is beyond the scope of this paper. Instead, this paper will present a method to generate the set of all consistent crime scenarios $\mathbf{S}$ and how it is used to constructed the set H_E of hypotheses supported by the available evidence.

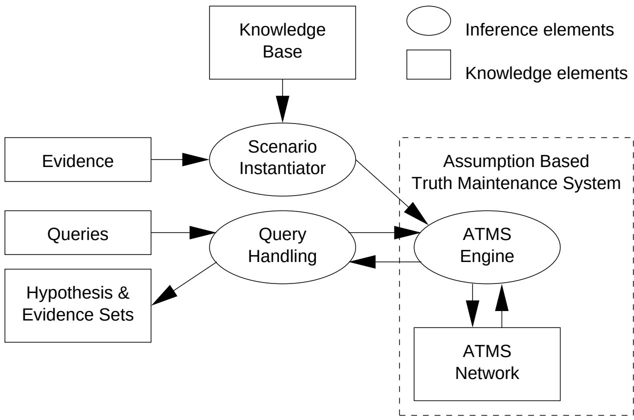

Figure 1: Basic architecture of the model based reasoner for crime investigation

4. CRIME SCENARIO ABDUCTION

Figure 1 shows the basic architecture of the proposed model based reasoning DSS. The central component of this architecture is an assumption based truth maintenance system (ATMS). An ATMS is an inference engine that enables a problem solver to reason about multiple possible worlds. Each possible world describes a specific set of circumstances, a crime scenario in this particular application, under which certain events and states are true and other events and states are false. What is true in one possible world, may be false in another. The task of the ATMS is to maintain what is true in each possible world. The way in which an ATMS works will be described in more detail in section 4.1

The ATMS is employed by two separate problem solvers. First, the *scenario instantiator* constructs the space of possible worlds. Given a knowledge base that contains a set of generic reusable components of a crime scenario and a set of pieces of evidence, the scenario instantiator builds a space of all the plausible crime scenarios, called the *scenario space*, that may have produced the complete set of pieces of evidence. The algorithm that implements the scenario instantiator is presented and illustrated in section 4.2.

Once the scenario space is constructed, it can be analysed by the *query handler*. The query handler can provide answers to the following questions:

- Which hypotheses are supported by the available evidence?
- What additional pieces of evidence can be found if a certain scenario/hypothesis is true?
- What pieces or sets of additional evidence can differentiate between two hypotheses?

This will be discussed in section 4.3.

4.1 Assumption Based Truth Maintenance

An ATMS is mechanism that maintains how each piece of inferred information depends on presumed information and facts and how inconsistencies arise. This section summarises the functionality of an ATMS as it is employed by this work. For more details, the reader is referred to the original papers [4, 5].

In an ATMS, each piece of information of relevance to the problem solver is stored as a *node*. Some pieces of information are not known to be true and cannot be inferred from other pieces of information. The plausibility of these is determined through the inferences made from them. In the ATMS, they are represented by a special type of node, called *assumption*.

Inferences between pieces of information are maintained within the ATMS as inferences between the corresponding nodes. In its extended form (see [5] or [13]), the ATMS can take inferences, called *justifications* of the form $n_i \wedge \ldots \wedge n_j \wedge \neg n_k \wedge \ldots \wedge \neg n_l \rightarrow n_m$, where $n_i, \ldots, n_j, n_k, \ldots, n_l, n_m$ are nodes (and assumptions) representing things that the problem solver is interested in. An ATMS can also take justifications, called *nogoods* that have lead to an inconsistency, i.e. justifications of the form $n_i \wedge \ldots \wedge n_j \wedge \neg n_k \wedge \ldots \wedge \neg n_l \rightarrow \perp$. The latter nogood implies that at least one of the statements in $\{n_i, \ldots, n_j, \neg n_k, \ldots, \neg n_l\}$ must be false.

Based on the given justifications and nogoods, the ATMS computes a *label* for each (non-assumption) node. A label is a set of *environments* and an environment is a set of assumptions. An environment E depicts a possible world where all the assumptions in E are true. The label $\mathcal{L}(n)$ of a node n describes all possible worlds in which n can be true. For reasons of efficiency and effectiveness, the label computation algorithm guarantees that each label is:

- *Sound*: Presuming that all assumptions in an environment from the label of a node are is true is a sufficient condition to derive that node. Formally, $\mathcal{L}(n)$ is sound if

$$\forall E \in \mathcal{L}(n), [(\wedge_{n_i \in E} n_i) \wedge (\wedge_{\neg n_i \in E} \neg n_i)] \vdash n$$

- *Consistent*: No environment in the label of a node describes an impossible world (i.e. a world from which $\perp$ logically follows). Formally, $\mathcal{L}(n)$ is consistent if

$$\forall E \in \mathcal{L}(n), [(\wedge_{n_i \in E} n_i) \wedge (\wedge_{\neg n_i \in E} \neg n_i)] \nvdash \perp$$

- *Complete*: The label describes all possible worlds. Formally, a label is complete if

$$\forall E, \exists E' \in \mathcal{L}(n),$$
$$[(\wedge_{n_i \in E} n_i) \wedge (\wedge_{\neg n_i \in E} \neg n_i) \vdash n] \rightarrow (E' \subseteq E)$$

- *Minimal*: The label does not contain possible worlds that are less general than one of the other possible worlds it contains (i.e. environments that are supersets of other environments in the label). Formally, a label is minimal if

$$\forall E \in \mathcal{L}(n) \nexists E',$$
$$[(\wedge_{n_i \in E'} n_i) \wedge (\wedge_{\neg n_i \in E'} \neg n_i) \vdash n] \wedge E' \subseteq E$$

4.2 Scenario Instantiation

This subsection presents a novel algorithm that has been devised for scenario generation. First, the knowledge representation formalism employed to construct the knowledge base is discussed, followed by a presentation of the actual algorithm that instantiates the knowledge base into a set of scenarios (the scenario space).

The theory is illustrated by examples taken from the domain of the differentiation between homicidal, suicidal, accidental and natural death. The case considered herein involves homicidal or accident death of babies due to a subdural haemorrhage. A subdural haemorrhage is a leakage of blood from vessels on the underside of the dura, one of the membranes covering the brain. It is a common cause of death of abused babies (the so-called shaken baby syndrome), but the injury may also be due to a number of non-homicidal causes, such as complications at birth, early childhood illnesses and certain medical procedures.

4.2.1 Knowledge Representation

In this work, it is presumed that the states and events constituting a scenario can be represented as predicates or relations. Although future work will expand on this basic representation formalism, this description is sufficient for dealing with many types of complex scenarios.

Naturally, states and events do not exist in isolation from one another. Certain states or events may be consequences of combinations of other states and events. For example, if a person is being assaulted and capable of self-defence, then (s)he will probably engage in some form of defensive action. Such knowledge is represented by means of *scenario fragments*.

Definition 1 (Scenario Fragment) *A* scenario fragment μ *is a tuple* $\langle V, V^s, V^t, \Phi^s, \Phi^t, A\rangle$ *where*

- $V = \{v_1, \ldots, v_l\}$, $V^s(\mu) = \{p^s_1, \ldots p^s_m\}$ *and* $V^t(\mu) = \{v^t_1, \ldots, v^t_n\}$ *are sets of variables,*
- $\Phi^s(\mu) = \{\phi^s_1, \ldots, \phi^s_v\}$ *is a set of relations, called* preconditions, *whose free variables are elements of* $V^s \cup V^t$,
- $\Phi^t(\mu) = \{\phi^t_1, \ldots, \phi^t_w\}$ *is a set of relations, called* postconditions, *whose free variables are elements of* P^t, *and*
- $A(\mu) = \{a_1, \ldots, a_t\}$ *is a set of relations, called* assumptions

such that for $i = 1, \ldots, w$*:*

$$\forall v_1, \ldots, \forall v_l, \forall v^s_1, \ldots, \forall v^s_m, \exists v^t_1, \ldots, \exists v^t_n\ \phi^s_1 \wedge \ldots \wedge \phi^s_v \rightarrow (a_1 \wedge \ldots \wedge a_t \rightarrow \phi^t_i)$$

To illustrate the concept of scenario fragment, consider the example below:

```
if {
  doctor(D), person(B)
  subdural-haemorrhage(B)
} assuming {
  cause-of-death(B,subdural-haemorrhage),
  correct-diagnosis(D,cause-of-death(B))
} then {
  medical-report(D,cause-of-death(B),
                 subdural-haemorrhage)
}
```

This scenario fragment states the following: given a person B, a doctor D and the fact that B suffered a subdural haemorrhage; and assuming that the cause of death of B is the subdural haemorrhage and that D makes a correct diagnosis of that cause of death; then a medical report must exist, written by D, stating that the cause of death of B is a subdural haemorrhage. It matches definition 1 as follows:

$$\begin{aligned}
V^s &= \{\} \\
V^t &= \{\} \\
V &= \{\texttt{B}, \texttt{D}\} \\
\Phi^s &= \{\texttt{doctor(D)}, \texttt{person(B)}, \\
&\quad \texttt{subdural-haemorrhage(B)}\} \\
A &= \{\texttt{cause-of-death(B,subdural-haemorrhage)}, \\
&\quad \texttt{correct-diagnosis(D,cause-of-death(B))} \\
\Phi^t &= \{\texttt{medical-report(D,cause-of-death(B),} \\
&\quad \texttt{subdural-haemorrhage)}\}
\end{aligned}$$

Some states and events are inconsistent with one another. For example, a terminal illness and a fatal injection of insulin cannot both be the cause of death of a person. Such knowledge is represented by means of *inconsistencies*.

Definition 2 (Inconsistency) *An inconsistency is a tuple* $\langle V, \Phi\rangle$ *where* $V = \{v_1, \ldots v_l\}$ *is a set of variables and* $\Phi = \{\phi_1, \ldots, \phi_v\}$ *is a set of relations, whose free variables are elements of* P*, such that:*

$$\forall v_1, \ldots, \forall v_l, (\wedge_{\phi_i \in \Phi} \phi_i) \rightarrow \perp$$

In this work, the states describing hypotheses and evidence are presumed to be consequences of other states and events. This presumptions quite naturally fits the original definition of section 3 quite naturally. Hypotheses are properties of scenarios, and hence, they should logically follow from a subset of the scenario. Similarly, pieces of evidence were defined as "observable consequences".

4.2.2 Inference Mechanism

The goal of the scenario instantiator is to construct a space of plausible crime scenario by instantiating the knowledge base of scenario fragments and inconsistencies into an ATMS. A novel algorithm GENERATESCENARIOSPACE($O, \mathbf{S}, \mathbf{I}$) where O is a set of evidence, $\mathbf{S}$ is a set of scenario fragments and $\mathbf{I}$ is a set of inconsistencies has been devised for this purpose. This algorithm expands on an existing composition modelling algorithm devised for the automated construction of ecological models [13]. As illustrated in figure 2, it consists of four phases:

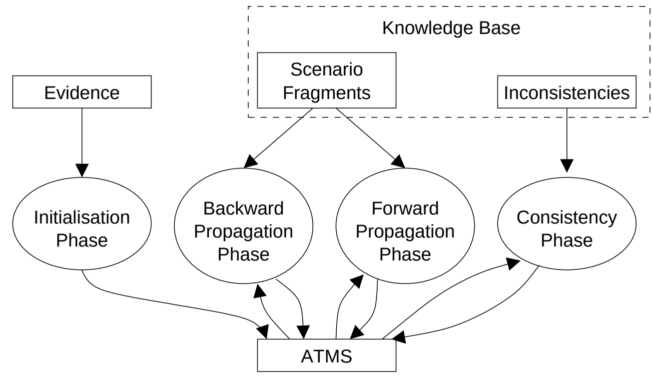

Figure 2: Overview of the inference mechanism

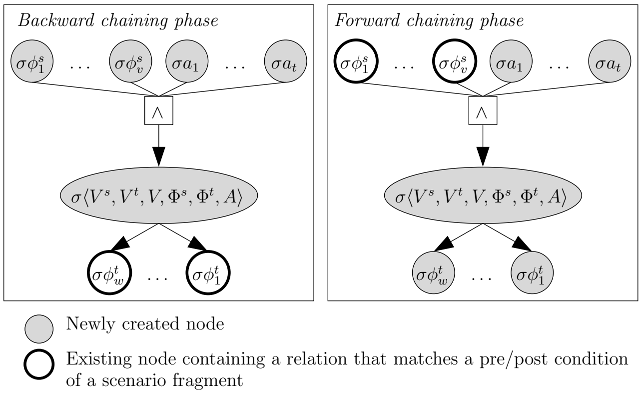

Figure 3: Scenario fragment instantiation in the ATMS

1. *Initialisation phase*: Here, an ATMS θ is created and initialised by adding a node for each given piece of evidence.

2. *Backward chaining phase*: All combinations of possible events and states that can possible produce the given pieces of available evidence O are reconstructed. That is, for each scenario fragment whose postconditions match relations in the ATMS θ (that is, the scenario fragments $\langle V, V^s, V^t, \Phi^s, \Phi^t, A\rangle$ for which a substitution σ exist that maps the postconditions Φ^t to relations referred to by nodes in the ATMS), a new set of nodes and justifications is added to the ATMS as follows:

   - A node m is added to θ denoting the application of the scenario fragment. Each node n to which the scenario fragment was matched (i.e. each node $\eta(\sigma\phi)$ with $\phi \in \Phi^t$) is justified in the ATMS by $n \leftarrow m$.
   - For each variable $v \in V^s$ a new constant c is created and the substitution $\{v/c\}$ is added to σ.
   - A node denoting $\sigma\phi$ is added for each precondition $\phi \in \Phi^s$ and an assumption denoting σa is added for each assumption $a \in A$. The conjunction of these newly created nodes is added as a justification of node m (the node denoting the instantiation of the scenario fragment).

   The resulting nodes and justifications are shown graphically in figure 3. Initially, θ is only populated with the pieces of evidence given in O and the algorithm works its way backwards to determine the potential sources of those pieces of evidence as described in the knowledge base of scenario fragments **S**.

3. *Forward chaining phase*: All the pieces of evidence and hypotheses that can be consequences of plausible scenarios generated in the backward chaining phase are extrapolated. For each scenario fragment whose preconditions match relations in the ATMS θ (i.e each scenario fragments $\langle V, V^s, V^t, \Phi^s, \Phi^t, A\rangle$ for which a substitution σ exist that maps the postconditions in Φ^s to relations referred to by nodes in the ATMS), a new set of nodes and justifications is added to the ATMS as follows:

   - For each variable $v \in V^t$ a new constant c is created and the substitution $\{v/c\}$ is added to σ.
   - A node denoting $\sigma\phi$ is added for each postcondition $\phi \in \Phi^t$ and an assumption denoting σa is added for each assumption $a \in A$.
   - A node m is added to θ denoting the application of the scenario fragment. This new node is justified by the conjunction of the instances of the relations in Φ^s and A:

     $$m \leftarrow [(\wedge_{\phi\in\Phi^s}\sigma\phi) \wedge (\wedge_{a\in A}\sigma a)]$$

     Each postcondition instance is justified by the new node m.

   The resulting nodes and justifications are also shown graphically in figure 3.

4. *Consistency phase*: In the final stage, the inconsistencies are instantiated and reported to the ATMS θ. More specifically, for each inconsistency whose relations match relations in the ATMS θ (that is, each inconsistency $\langle V, \Phi\rangle$ for which a substitution σ exist that maps the relations in Φ to relations referred to by nodes in the ATMS), a nogood $\bot \leftarrow (\wedge\phi \in \Phi\sigma\phi)$ is created.

A formal representation of the algorithm is given below:

Algorithm 4.1: GENERATESCENARIOSPACE(O, **S**, **I**)

comment: *Initialisation phase*:
$\theta \leftarrow$ new ATMS;
for each $e \in O$, add-node(θ, e);
comment: *Backward chaining phase*:
for each $\langle V, V^s, V^t, \Phi^s, \Phi^t, A\rangle \in \mathbf{S}, \exists\sigma, \text{match}(\langle V \cup V^t, \Phi^t\rangle, \theta, \sigma)$
do
  $J \leftarrow \emptyset$;
  for each $v \in V^s, \sigma \leftarrow \sigma \cup \{v/\text{gensym}()\}$;
  for each $\phi \in \Phi^s$
    do $n \leftarrow$ ADDNODE$(()\theta, (\sigma\phi))$;
      $J \leftarrow J \cup \{n\}$;
  for each $a \in A$
    do $n \leftarrow$ ADDASSUMPTION$(()\theta, (\sigma a))$;
      $J \leftarrow J \cup \{n\}$;
  $m \leftarrow$ ADDNODE$(()\theta, (\sigma\langle V, V^s, V^t, \Phi^s, \Phi^t, A\rangle))$;
  ADDJUSTIFICATION$(()\theta, m, J)$;
  for each $\phi \in \Phi^t$, ADDJUSTIFICATION$(()\theta, \eta(\sigma\phi), \{m\})$;
comment: *Forward chaining phase*:
for each $\langle V, V^s, V^t, \Phi^s, \Phi^t, A\rangle \in \mathbf{S}, \exists\sigma, \text{match}(\langle V \cup V^s, \Phi^s\rangle, \theta, \sigma)$
do
  $J \leftarrow \emptyset$;
  for each $\phi \in \Phi^s, J \leftarrow J \cup \{(\sigma\phi)\}$;
  for each $a \in A$
    do $n \leftarrow$ ADDASSUMPTION$(()\theta, (\sigma a))$;
      $J \leftarrow J \cup \{n\}$;
  $m \leftarrow$ ADDNODE$(()\theta, (\sigma\langle V, V^s, V^t, \Phi^s, \Phi^t, A\rangle))$;
  ADDJUSTIFICATION$(()\theta, m, J)$;
  for each $v \in V^t, \sigma \leftarrow \sigma \cup \{v/\text{gensym}()\}$;
  for each $\phi \in \Phi^t$
    do $n \leftarrow$ ADDNODE$(()\theta, (\sigma\phi))$;
      ADDJUSTIFICATION$(()\theta, n, \{m\})$;
comment: *Consistency phase*:
for each $\langle V, \Phi\rangle \in \mathbf{I}, \exists\sigma$, MATCH$(()\langle V, \Phi\rangle, \theta, \sigma)$
do
  $J \leftarrow \emptyset$;
  for each $\phi \in \Phi, J \leftarrow J \cup \{(\sigma\phi)\}$;
  ADDNOGOOD$(()\theta, J)$;

The scenario instantiation algorithm employs a function MATCH$(\langle V, \Phi\rangle, \theta, \sigma)$ to find instances of the relations in the scenario fragments and inconsistencies. The function takes the following arguments: 1) a set of free variables V, 2) a set of relations Φ whose free variables are elements of V, 3) the ATMS under construction θ and a substitution σ. The substitution σ maps each variable in $v \in V$ to a constant σv

and each relation $\phi \in \Phi$ to a grounded relation $\sigma\phi$ where the variables are substituted by constants.

The function is true if for each relation $\phi \in \Phi$, a node exists in the ATMS θ that denotes the grounded relation $\sigma\phi$. Formally, $\textsc{match}(\langle V, \Phi \rangle, \theta, \sigma)$ is deemed true if

$$(V = \{v_1, \ldots, v_p\}) \wedge (\sigma = \{v_1/o_1, \ldots, v_p/o_p) \wedge$$
$$(\forall \phi_j \in \Phi, \eta(\sigma\phi_j) \in \theta)$$

where η is the function that maps grounded relations of interest to our problem solver (the crime scenario instantiator) to nodes and assumptions in ATMS.

An example of a partial scenario space, in the sample application domain, that can be constructed in this way is presented in figure 4. This scenario space contains 13 instances of scenario fragments computed based on two pieces of evidence: 1) a medical report stating that the cause of death of a baby b is a subdural haemorrhage and 2) the observation that the child has bruises on his/her body.

The top half of the figure depicts the scenario fragments that have been instantiated in the backward chaining phase. During this phase, all possible causes (as far as they are included in the knowledge base) of the two pieces of evidence are generated. In this case, two sets of events and states (i.e. scenarios) can be distinguished. In the first scenario, baby b died due to a subdural haemorrhage caused by abuse by a person p. In the second scenario, baby b died due to inadequate collagen synthesis. The lack of collagen could have caused a weakened blood vessel, which in turn led to a subdural haemorrhage, as well as weakened bones, which explains the bruises.

The bottom half of the figure contains the scenario fragments that have been instantiated in the forward chaining phase. During this phase, hypotheses and additional sources of evidence are generated. In this case, the first scenario (the one suggesting abuse) would prove the homicide hypothesis and the second scenario describes an accidental death. Potential additional sources of evidence are an exploration of bruises or an examination of the collagen synthesis function in the baby.

Note that the scenario space accommodates for the possibility that certain pieces of evidence may be misleading. Indeed, most pieces of evidence are dependent upon an assumption denoting the presumption that no mistakes were made in the interpretations or observations that lead to the evidence. These assumptions allow for the possibility of alternative explanations, say, in case the evidence is seemingly inconsistent.

4.3 Decision Making

The ATMS θ constructed by the algorithm described in section 4.2.2 contains a space of all scenarios that can be constructed with the knowledge base and that produce the given set of evidence O. This section shows how the information contained in this ATMS can be exploited to answer the aforementioned three types of query. The approach taken herein involves translating queries into formal ATMS nodes and justifications, thus enabling the existing ATMS label propagation to answer the queries of interest.

Formally, any consistent conjunction of assumptions that entails all pieces of evidence (in θ) constitutes a possible world for the case under investigation. Any set of assumption W such that

$$[(\wedge_{a \in W} a), \theta \nvdash \perp] \wedge [\forall e \in O, (\wedge_{a \in W} a), \theta \vdash e]$$

and the consequences of those assumptions describe a plausible crime scenario. Therefore, all the scenarios contained in the ATMS can retrieved by computing the label for the conjunction of the pieces of evidence in O. Let n_O be an additional node added to θ and justified as:

$$n_O \leftarrow (\wedge_{e \in O} \eta(e))$$

Then, the label $\mathcal{L}(n_O)$ contains all the environments from which plausible scenarios can be produced. This knowledge enables the decision support system to answer the following questions:

- *Which hypotheses are supported by the available evidence?* Every hypothesis that follows from a plausible scenario is supported by the available evidence. That is, a hypotheses h is supported by the evidence if it follows from an environment of the label of n_O:

  $$\exists W \in \mathcal{L}(n_O), (\wedge_{a \in W} \eta(a)), \theta \vdash \eta(h)$$

  where $\eta(e)$ refers to the node that denotes e in θ. Thus, if the label of h is not empty, then that hypothesis is supported by the available evidence.

  It the scenario space of figure 4, there are two environments that support the available (two pieces of) evidence:

  $O_1 = \{$lack-of-collagen(b),
  accidental-subdural-blood-vessel-rupture(b),
  cause-of-death(b,subdural-haemorrhage),
  correct-diagnosis(d,cause-of-death(b))$\}$

  $O_2 = \{$abuse(p,b),
  subdural-blood-vessel-rupture-due-to-abuse(b),
  cause-of-death(b,subdural-haemorrhage),
  correct-diagnosis(d,cause-of-death(b))$\}$

  In the possible world described by environment O_1, accident(b) is true and in the one described by O_2, homicide(b) is true. Therefore, it follows that both hypotheses are supported by the available evidence.

- *What additional pieces of evidence can be found if a certain scenario/hypothesis is true?* All the states and events, including pieces of evidence, that are logical consequence states and events in plausible scenarios are generated in the forward chaining phase of the algorithm. Therefore, θ will contain nodes representing pieces of evidence that are produced in certain scenarios but were not collected in O. As with the hypotheses, the labels of these nodes describe the environments (and hence, the scenarios) under which these pieces of evidence are expected.

  A piece of evidence e can be found under a given hypothesis h if a possible world exists that supports both the evidence and the hypothesis. This is the case if an environment in the label of the evidence entails an environment in the label of the hypothesis:

  $$\exists W_h \in (\eta(h)), \exists W_e \in (\eta(e)), W_h \subseteq W_e$$

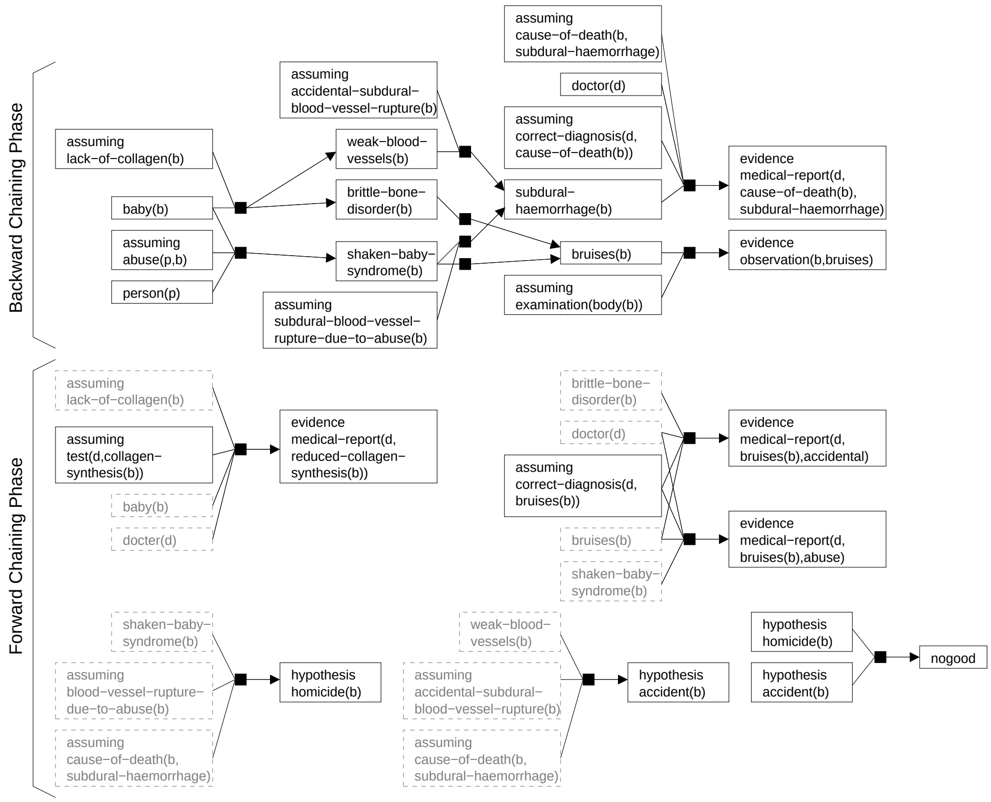

Figure 4: Sample partial scenario space

Continuing with the ongoing example, in the scenario space of figure 4, a piece of evidence e that consists of a medical report documenting reduced collagen synthesis in b, medical-report(d, reduced-collagen-synthesis(b)), is generated under the environment:

$$
\begin{aligned}
O_3 = \{&\text{lack-of-collagen(b)},\\
&\text{accidental-subdural-blood-vessel-rupture(b)},\\
&\text{cause-of-death(b,subdural-haemorrhage)},\\
&\text{correct-diagnosis(d,cause-of-death(b))},\\
&\text{test(d,test-collagen-synthesis(b))}\}
\end{aligned}
$$

Because $O_1 \subset O_3$, it follows that under the accident(b) hypothesis, evidence e may be found.

- *What pieces or sets of additional evidence can differentiate between two hypotheses?* Let h_1 and h_2 be two hypotheses, then any set of pieces of evidence O' that can be found if h_1 is true, but are inconsistent with h_2 can differentiate between the two hypotheses. As reported in [15], this idea will be extended in future work to compute the degree to which evidence differentiates between the relative probability of hypotheses. For example, it follows from the above discussion that the piece of evidence medical-report(d, reduced-collagen-synthesis(b)) may help to differentiate between the two hypotheses, accident(b) and homicide(b). This information suggests to a social worker or police officer examining the case that ordering a test for reduced collagen synthesis would be useful.

5. DISCUSSION AND FUTURE WORK

The main advantage of the approach presented herein is its robustness. The scenario space generation algorithm can compose combinations of events and states that produce a given set evidence from a knowledge base of generic scenario fragments and inconsistencies. Therefore, the crime scenarios that the system will be confronted with need not be anticipated during the knowledge acquisition phase.

Of course, crime investigation DSSs employing the approach presented herein still require a significant knowledge acquisition effort to construct a knowledge base describing how the events and states are related to one another and which sets of events and states are inconsistent. But, it can be argued that the events and states constituting the scenarios recur much more frequently than the scenarios themselves. For example, there are a finite number of causes a subdural haemorrhage, such as a trivial fall or a blow to the head. However, such injuries can occur in a wide variety of circumstances ranging from child abuse to alcoholism. Each of these circumstances can also be described by a set of events and states that are not restricted to causes of subdural haemorrhages. Therefore, these events and states are far more common than the specific scenarios in which they occur.

The approach presented herein aids crime investigators by reasoning about multiple scenarios simultaneously. Generally speaking, humans tend to focus on one or a small number of scenarios and they search for evidence that confirms a single scenario or differentiates between a few scenarios. This can be a problem in complex cases because miscarriages are often a consequence of ignoring a plausible scenario and inadequate evidence collection to differentiate between plausible scenarios. The DSS presented herein aims to alleviate these issues by complementing the skills of the human investigator.

For the DSS to work optimally, the set of scenario fragments $\mathbf{S}$ and the set of inconsistencies $\mathbf{I}$ that are employed to construct θ should be both sound and complete. The knowledge bases are said to be *sound* if every scenario fragment describes a real-world cause and effect relation between events and states and if every inconsistency contains a set of states and events that is physically impossible. The knowledge bases are said to be *complete* if every possible real-world cause and effect relation between events and states and every source of inconsistency is represented in the knowledge bases. Obviously, it is very difficult to achieve completeness in practice since the knowledge of any group of human experts on crime investigation, from which the knowledge bases are constructed, is inevitably incomplete. However, if the DSS employs a knowledge base that is more complete than the knowledge of the user, then the DSS is able to consider scenarios that human investigators overlook.

In future work, the method presented herein will be expanded upon. Firstly, the representation formalisms employed to describe states and events in crime scenarios will be elaborated. As described earlier, the sets of states and events that constitute a scenario are restricted by the consistency requirements. This paper introduced a generic means to represent when inconsistencies occur and to prevent inconsistent scenarios from being considered when hypotheses are generated and evidence collection strategies are constructed. When reasoning about related events that take place over time and space, temporal and spatial constraints are an important source of such inconsistencies. To avoid overcomplicating this paper, the important issues of temporal and spatial reasoning were not considered, but will be addressed in future work.

Secondly, methods are under development to assess the relative likelihoods of alternative scenarios. In [15], several methods to expand the entropy based decision making techniques employed by model based diagnosis techniques [9] were presented. The application of these methods requires a means of generating a space of alternative scenarios and a way of computing the relative probabilities of alternative scenarios. The former has been introduced in this paper and the latter will be presented in a future paper.

Thirdly, an extensive knowledge base will be developed to enable the deployment of this system. Currently, a prototype implementing the algorithms described herein has been developed. This has enabled the validation of the theory and the example used in this paper. However, it is clear that a proper evaluation of the approach requires its application to a real-world domain problem.

6. CONCLUSIONS

In this paper, an algorithm has been presented that is capable of automatically generating a space of crime scenarios that can explain a given set of available evidence. It employs a knowledge base of formal descriptions of component events that recur in different combinations in crime scenarios, and it instantiates and composes them to form relevant instances of such scenarios. The main advantage of this approach lies in its robustness: because there are far fewer component events than possible combinations of such events, this system is more likely to be able to handle previously unseen cases than an similar system that employs a knowledge base of full scenario descriptions.

The paper has also discussed how an automatically constructed space of plausible crime scenarios can help crime investigators to formulate hypotheses about the case at hand and help them in the collection of further evidence. As such, this work is a first step in the development of a decision support system for crime investigation. Yet, it is recognised that, despite the robustness of the proposed approach, any formalisation of a large body of real-world knowledge such a plausible crime scenarios is prohibitively expensive. Therefore, future applications of this work will focus on complex, but suitably restricted subproblems in major crime investigation, such as the differentiation between homicidal, suicidal, accidental and natural deaths in well-defined circumstances.

Acknowledgements

Both authors are grateful to their fellow members of the Joseph Bell Centre for Forensic Statistics and Legal Reasoning, whilst taking full responsibility for the views and ideas presented in this paper.

Prototype Software

The ideas presented in this paper have been successfully implemented in a small prototype system. This system is freely available for download at:

http://homepages.inf.ed.ac.uk/jeroen/mbdss/download.html

7. REFERENCES

[1] Badiru, A.B., Karasz, J.M., and Holloway, R.T. Arest: Armed robbery eidetic suspect typing expert system. *Journal of Police Science and Administration*, 16(3):210–216, 1988.

[2] Brown, D.E. The regional crime analysis program (RECAP): A framework for mining data to catch criminals. In *Proceedings for the 1998 International Conference on Systems, MAN, and Cybernetics*, pages 2848–2853, 1998.

[3] Chen, H., Schroeder, J, Hauck, R.V., Ridgeway, L., Atabakhsh, H., Gupta, H., Boarman, C., Rasmussen, K., and Clements, A.W. COPLINK Connect: Information and knowledge management for law enforcement. *Decision Support Systems: Special Issue on Digital Government*, 34(3):271–285, 2002.

[4] de Kleer, J. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.

[5] de Kleer, J. A general labeling algorithm for assumption-based truth maintenance. In *Proceedings of the 7th National Conference on Artificial Intelligence*, pages 188–192, 1988.

[6] Dzemydiene, D. and Rudzkiene, V. Multiple regression analysis in crime pattern warehouse for decision support. In *Proceedings of the 13th International Conference on Database and Expert Systems Applications*, pages 249–258, 2000.

[7] Esmaili, M., Safavi-Naini, R., Balachandran, B., and Pieprzyk, J. Case-based reasoning for intrusion

detection. In *IEEE Twelfth Annual Computer Security Applications Conference*, pages 214–223, 1996.

[8] Falkenhainer, B. and Forbus, K. Compositional modeling: finding the right model for the job. *Artificial Intelligence*, 51:95–143, 1991.

[9] Hamscher, W., Console, L., and de Kleer, J., editors. *Readings in Model-Based Diagnosis*. Morgan-Kaufmann, San Mateo, 1992.

[10] Hand, D., Mannila, H., and Smyth, P. *Principles of Data Mining*. MIT Press, Boston, Massachusetts, 2001.

[11] Hauck, R.V., Atabakhsh, H., Ongvasith, P., Gupta, H., and Chen, H. Using coplink to analyze criminal justice data. *IEEE Computer*, 35(3):30–37, 2002.

[12] Jung, C., Han, I., and Suh, B. Risk analysis for electronic commerce using case-based reasoning. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 8:61–73, 1999.

[13] Keppens, J. *Compositional Ecological Modelling via Dynamic Constraint Satisfaction with Order-of-Magnitude Preferences*. PhD thesis, The University of Edinburgh, 2002.

[14] Keppens, J. and Shen, Q. On compositional modelling. *Knowledge Engineering Review*, 16(2):157–200, 2001.

[15] Keppens, J. and Zeleznikow, J. On the role of model-based reasoning in decision support in crime investigation. In *Proceedings of the 3rd International Conference on Law and Technology*, pages 77–83, 2002.

[16] Prakken, H. Modelling reasoning about evidence in legal procedure. In *Proceedings of the 8th Internation Conference on Artificial Intelligence and Law*, pages 119–128, 2001.

[17] Prakken, H., Reed, C., and Walton, D. Argumentation schemes and generalisations in reasoning about evidence. In *Proceedings of the 9th Internation Conference on Artificial Intelligence and Law*, 2003.

[18] Prakken, H. and Sartor, G. A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law*, 4:331–368, 1996.

[19] Ratledge, E.C. and Jacoby, J.E. *Handbook on Artificial Intelligence and Expert Systems in Law Enforcement*. Greenwood Press, Westport, Connecticut, 1989.

[20] Ribaux, O. and Margot, P. Inference structures for crime analysis and intelligence: the example of burglary using forensic science data. *Forensic Science International*, 100:193–210, 1999.

[21] Schum, D. *The Evidential Foundations of Probabilistic Reasoning*. Northwestern University Press, 1994.

[22] Valcour, L. Investigate B& E: Break & enter expert system. Technical Report TR-11-97, Canadian Police Research Centre, 1997.

[23] Walton, D. Argumentation and theory of evidence. *New Trends in Criminal Investigation and Evidence*, 2:711–732, 2000.