A logic-based approach to understanding lone-actor terrorism

Dalal Alrajeh
Department of Computing, Imperial College London
London SW7 2RH, United Kingdom
dalal.alrajeh@imperial.ac.uk

Paul Gill
Department of Security and Crime Science, University College London
London WC1H 9EZ, United Kingdom
paul.gill@ucl.ac.uk

submitted 29 April 2015; accepted 5 June 2015

Abstract

The need for systematic research into behavioural factors of individual terrorists has been highlighted by much recent work on terrorism. Many existing methods follow a hypothesis-testing approach in which statistical modelling and analysis of existing data is conducted to either confirm or refute a hypothesis. However, the initial construction of hypotheses is not trivial, nor is the decision upon which of the variables are to be considered relevant for the testings. It has been argued that the lack of a methodical approach to represent, analyse, interpret and infer from existing data presents a pressing challenge to the progress of lone-actor terrorism research in particular, and the terrorism field more generally.

This paper sets a new agenda for such research. We propose the use of a logic programming approach to address the shortcomings of existing methodologies in the study of lone-actor terrorism. Our method is based on transforming characteristic and behavioural codes into a logic program and applying inductive logic programming to learn hypotheses about potentially relevant factors associated with terrorist behaviour, as well as the influence of specific factors on such behaviour. This paper is an exploratory study of 111 lone-actor terrorists' target selections (civilian vs. high-value targets) and the agency of their ideological orientation in determining their target choices.

1 Introduction

An emerging consensus within terrorism studies posits analysing what terrorists do as opposed to merely studying who they are is more instructive (Horgan 2014). A growth in datasets focused upon individual, as opposed to group, behaviour has fostered a major new development in our understanding of terrorist behaviour (LaFree 2013). Rather than employing a single conception of the 'terrorist', these analyses disaggregate the sample and compare the subsets across specific characteristics and behaviours (Gill and Corner 2013). However, a number of problems remain endemic within the study of the individual terrorist. First, most of these new analyses rely upon testing hypotheses derived from the study of general criminal offenders whose psychologies and decision-making repertoires may not be generalizable to violent, politically-oriented offenders. In a field as underdeveloped as terrorism studies, hypothesis generation based purely on studies of terrorist behaviour can be onerous. Second, commonly used statistical methods (bivariate and multivariate analyses) fail to capture causality relations between variables. These approaches also concentrate on subsets of variables theoretically linked with the observation being investigated. However, in some cases, this is not sufficient as they may only be explained by combinations of loosely related variables that potentially remain untested. Statistical findings do not speak for themselves: moving from 'factors' to policy is not straightforward (Farrington 2000). As a consequence, it has been argued that a knowledge-base capable of supporting policy must contain more than a catalogue of factors, however exhaustive. It must include theories which advance explanations of how these factors are related to the outcome of interest (Wikstrom 2011). Knowledge is achieved when outcomes are explained, rather than merely described, or even predicted.

In this paper, we explore a logic programming approach to representing and reasoning about lone-actor terrorists' characteristics and behaviours. We propose an approach for automatically generating hypotheses about lone actor terrorists with the aim of gaining better understanding of the links between individuals' characteristics and behaviour with respect to the outcome of their event planning. In particular, we investigate the use of inductive logic programming (ILP) to conduct two types of analyses: (i) identifying factors that are associated with and can differentiate between terrorists' target selections (i.e., *learning associations*), and (ii) capturing the influence of specific factors in explaining such differences (i.e., *learning influences*). The approach is applied to a dataset containing antecedent behaviours and characteristics of 111 lone actors terrorists, originally described in (Gill et al. 2014). The overall objective of this work is to provide an exploratory approach that overcomes limitations of the methods standardly applied in this area of study by: (a) automatically generating explanations that are guaranteed to cover all the observations; (b) suggesting alternative hypotheses to be tested; and (c) demonstrating how the *presence* or *absence* of a single factor can be associated with different observed outcomes when combined with others. The approach presented in this paper forms the initial steps towards developing a logic-based, causal framework for reasoning about criminal behaviour with wider applicability in crime science studies.

2 Background

2.1 Lone-actor Terrorism Characteristic and Behavioural Codes

In this paper, *terrorism* is defined as a violent action, or threat of violent action, aimed at intimidating and coercing a government or sections of the public, typically for political, religious or ideological ends. Terrorism can involve violence against a person, damage to property, endangering a person's life other than that of the terrorist, creating a serious risk to the health or safety of the public or a section of the public. A *lone-actor* is either a single terrorist or an isolated dyad (a pair of individuals), operating independently of a group. *Antecedent behaviour* is the behaviour of the offenders leading up to their planning or conducting a terrorist act. *Demographic characteristics* of lone-actor terrorists include gender, education and socio-economic indicators such as employment status.

For analysis purposes, antecedent behaviour and demographic characteristics are typically represented as codes (also called *variables*). Each code c has a domain, denoted $dom(c)$, of possible values. When C is a set of codes, we write $dom(C)$ as a shorthand for the set of domains for each $c \in C$. We are interested here in codes whose domains are non-empty, finite sets of discrete values. A code is binary if its domain contains two values only, and multi-valued if it contains more than two. *Boolean* codes are binary codes with domain $\{true,false\}$. A code can be assigned one or more values from its domain. Given a set of codes C, the function $singlevalue : C \rightarrow \{true, false\}$ returns true if a code takes one value only and false otherwise. Boolean codes are single valued. Given a set ID containing a unique identifier for each lone-actor, we define a labelling function $\theta_c : ID \rightarrow 2^{dom(c)}$ for each code $c \in C$. Where $id \in ID$ is a lone-actor's identifier, $\theta_c(id)$ is a set of values from $dom(c)$, giving the actual values of *id*'s c code. We require that $\theta_c(id) \neq \emptyset$ in all cases and $\forall c : C \cdot (singlevalue(c) \rightarrow \forall id \in ID \cdot |\theta_c(id)| = 1)$. For example, the terrorist with identifier *pg011* targets both civilian and high-value groups: $\theta_{targetgroup}(pg011) = \{civilian, hvt\}$. We use θ_C to denote the set of assignment functions for each code in C. We sometimes refer to code assignments as factors. The full set of codes used in this paper is available at *www.doc.ic.ac.uk/~da04/iclp15/*.

2.2 Inductive Logic Programming

Inductive logic programming (ILP) (Muggleton 1991) is a logic-based machine learning technique the aims at automatically generating hypotheses from observed phenomena and a background theory expressed as logic programs. In this paper, we are concerned with learning normal logic programs. A normal logic program is one in which the clauses (or rules) are of the form $A \leftarrow B_1, \ldots, B_n, \textit{not } C_1, \ldots, \textit{not } C_m$ where A is the *head atom*, B_i are *positive body literals*, and *not* C_j are *negative body literals*. Normal logic programs may have one, none, or several (minimal) models. The semantics of our logic programs are based on *stable models* semantics (Gelfond and Lifschitz 1988). Given a normal logic program Π, the *reduct* of Π with respect to I, denoted Π^I, is the program obtained from the ground instances of Π by (a) removing all clauses with a negative literal *not* a in its body where $a \in I$ and (b) removing all negative literals from the bodies of the remaining clauses. If I is the least Herbrand model of Π^I then I is said to be a stable model of Π. This along with the notion of entailment are given below.

Definition 1
A model I of Π is a *stable model* if I is the least Herbrand model of Π^I where Π^I is the definite program $\Pi^I = \{A \leftarrow B_1, \ldots, B_n \mid A \leftarrow B_1, \ldots, B_n, \textit{ not } C_1, \ldots, \textit{ not } C_n$ is the ground instance of a clause in Π and I does not satisfy any of the $C_i\}$.

Definition 2
A logic program Π *entails* an expression ϕ (under the credulous stable model semantics), denoted $\Pi \models \phi$, iff ϕ is satisfied in at least one stable model of Π.

In ILP, mode declarations are used as a form of language bias to reduce the hypotheses search space. They provide a mechanism for specifying which predicates may appear in the heads and bodies of rules and for controlling the placement and linking of constants and variables within those clauses. A mode declaration M is either a head or body declaration, respectively *modeh*(s) and *modeb*(s) where s is called a schema. A schema s is a ground literal containing placemarkers. A placemarker is either '+type' (input), '−type' (output), '#type' (ground) where type is a constant. Given the above, an ILP task is defined as follows where $\models$ is interpreted under brave induction (Sakama and Inoue 2009).

Definition 3
A nonmonotonic ILP task is a tuple $\langle B, E^+, E^-, M\rangle$ where E^+ and E^- are sets of ground literals, called *positive examples* and *negative examples* respectively, B is a normal logic program, called *background theory* and M is a set of mode declarations defining a hypothesis space $s(M)$. An inductive solution (or hypothesis), $H \subseteq s(M)$, for $E^+ \cup E^-$ w.r.t B is a set of clauses such that:

$$B \cup H \models e^+, \forall e^+ \in E^+ \quad \text{and} \quad B \cup H \not\models e^-, \forall e^- \in E^-$$

under brave induction (Sakama and Inoue 2009).

We sometimes write $B \cup H \models E$ where $E = E^+ \cup E^-$ as a shorthand for the two conditions above.

In this paper, we focus on the use of a learning technique first introduced in (Ray 2009) and its implementation XHAIL. The technique is based on a three-phase Hybrid Abductive Inductive Learning (HAIL) approach (Ray et al. 2004). The XHAIL language and search bias mechanisms are based upon a compression heuristic that favours solutions containing the fewest number of literals.

3 Approach

We first introduce an automated mechanism for mapping characteristic and behavioural codes into a logic program. We then describe an ILP approach for conducting two types of analyses. The first aims at identifying combinations of factors that are associated with specific observations. We call this *learning associations.* The second is to understand the influence of specific factors when explaining these observations. We refer to this type as *learning influences.* Our focus in this paper will be on explaining terrorists' target group selection, civilian targets vs. high-value targets, where the latter includes targets such as government, business, schools etc. We demonstrate the approach using a dataset of 111 individuals (of which 91 have known target selections) and 185 codes described in (Gill et al. 2014).

3.1 Modelling Codes in Logic Programs

Our representation uses a sort *ID* for capturing the domain of unique lone-actor terrorists. It includes a unique predicate for each code c and type predicate to represent the domain of c. The number of arguments for this predicate depends on the size of the code's domain. If it is Boolean, e.g., *mentalill*, then, as a simplification, a predicate with a single argument of sort *ID* is introduced e.g., *mentalill(id).* If it is a non-Boolean variable, then a two argument predicate is introduced in which the first argument is of sort ID and the second is of sort $Dom(c)$. For instance, the nature of the location at which the attack occurred is encoded using a predicate $location\_nature \subseteq ID \times dom(location\_nature)$ where $dom(location\_nature) = \{government, business, private\_citizens, religious, military, other\}$. The language also contains a type predicate for representing $dom(c)$. The encoding is given below.

Definition 4
Let ID be a set of unique identifiers for lone-actors, C a set of codes named $\{c^1, \cdots, c^n\}$, $D(C)$ the set of C domains named $\{dc^1, \cdots, dc^n\}$ respectively and θ_C their assignment functions. The logic program $ID_{LP} \cup C_{LP} \cup D_{LP}(C) \cup A_{LP}$ encoding of ID, C, $D(C)$ and θ_C is constructed such that:

- ID_{LP} contains a fact $la\_iden(id)$ for each $id \in ID$;
- $D_{LP}(C)$ contains a fact $dc^k(j)$ for each $j \in dom(c^k)$;
- for each $id \in ID$ and $c \in C$, C_{LP} contains a fact $c^k(id)$ where c^k is a Boolean code and $\theta_c(id) = \{true\}$;
- for each $id \in ID$ and $c \in C$, C_{LP} contains a fact $c^k(id, j)$ where $|dom(c^k)| \geq 2$, c^k is non-Boolean, $j \in \theta_c(id)$ and $j \notin \{false, unknown\}$;
- for each $c \in C$, if c is non-Boolean and $singlevalue(c) = true$, A_{LP} contains the clause $\leftarrow c(I, D_1), c(I, D_2), D_1 \neq D_2\cdot$

Our encoding deploys a closed world assumption where unknown code values are treated as false, as assumed in (Gill et al. 2014). An example of the encoding is:

$$\begin{array}{lcl}
ID_{LP} & = & \{la_iden(pg018).\ la_iden(pgpg101)...\} \\
D_{LP}(C) & = & \{ideo(rightwing).\ ideo(single_issue)...\ loc(government)... \\
 & & tg(hvt).\ tg(citizens)...\} \\
C_{LP} & = & \{imprisoned(pg018).\ ideology(pg018,\ rightwing).\ mentalill(pg018). \\
 & & location_nature(pg018,\ government). \\
 & & ...\ dryruns(pg101).\ ideology(pg101,\ single_issue).\ f2f(pg101)...\} \\
A_{LP} & = & \{\leftarrow ideology(I, D_1), ideology(I, D_2), D_1 \neq D_2\cdot, \cdots\}
\end{array}$$

The expressiveness of the formalism allows us to capture relationships between codes. For example, a code value *both* may be introduced for capturing individuals who have targeted both civilian and high-value groups, in which case $dom(targetgroup)$ is extended with the value *both* and A_{LP} is amended with the following:

$$\begin{array}{l}
\{targetgroup(I, both) \leftarrow targetgroup(I, hvt), targetgroup(I, civilian)\cdot, \\
\leftarrow targetgroup(I, both), not\ targetgroup(I, hvt)\cdot, \\
\leftarrow targetgroup(I, both), not\ targetgroup(I, civilian)\cdot\}
\end{array}$$

3.2 Learning Characteristic and Behavioural Associations

In the context of characteristic and behavioural analyses, a code c_1 is said to be *associated with* another code $c2$ if $c1$ forms part of at least one explanation of the observed behaviour represented by $c2$. In ILP terms, identifying codes associated with an observation amounts to finding an inductive solution in which the literals corresponding to these codes appear in the body of at least one rule in the solution. In the presented approach, we do not distinguish between codes that are causes and those that are correlates. Future work will clarify the distinction.
Definition 5
Let ID be a set of unique lone-actor identifiers, C a set of codes, and $D(C)$ a set of C domains. Let $c_t \in C$ be a code representing the observed terrorist behaviour to

be explained, and $R = \{c_r \,|\, c_r \in C, c_r \neq c_t\}$ be the rest of variables in the language. The ILP task is defined with:

- $B = ID_{LP} \cup R_{LP} \cup D_{LP}(R) \cup D_{LP}(\{c_t\}) \cup A_{LP}$;
- E^+ includes a fact $c_t(id)$ for each $id \in ID$ where c_t is Boolean and $\theta_{c_t}(id) = \{true\}$, or a fact $c_t(id, j)$ where c_t is non-Boolean and $j \in \theta_{c_t}(id)$;
- E^- includes a fact $c_t(id)$ for each $id \in ID$ where c_t is Boolean and $\theta_{c_t}(id) = \{false\}$, or a fact $c_t(id, j)$ where c_t is non-Boolean, for every $j \in dom(c_t)$ such that $j \notin \theta_{c_t}(id)$;
- M includes $modeh(c_t(+la\_iden))$ (or $modeh(c_t(+la\_iden, \#dc_t))$), and a pair of body declarations $modeh(c_r(+la\_iden))$ (or $modeh(c_r(+la\_iden, \#dc_r))$) and $modeb(not\ c_r(+la\_iden))$ (or $modeb(not\ c_r(+la\_iden, \#dc_r))$) for each $c_r \in R$.

Note that in this type of the analysis, we do not impose any restrictions on the number of rules, within an inductive solution, in which a body literal appears to be said associated with the observation, nor on the number of observations explained by the rule in which it appear. However, to quantify the relevance of that association for a given dataset, our algorithm calculates a measure (*relative significance* value) for each hypothesis in an inductive solution as defined below.

Definition 6
Let B be a background theory, $E = E^+ \cup E^-$ the set of examples, M the mode declaration and $H = \{h_1, \cdots h_n\} \subseteq s(M)$ an inductive solution to E w.r.t B. Let $E^+_{h_i} \subseteq E^+$ be the set of positive examples explained by the hypothesis $h_i \in H$ such that $B \wedge h_i \models e^+_{h_i}$, for each $e^+_{h_i} \in E^+_{h_i}$. Then the relative significance σ of h_i, denoted $\sigma(h_i)$, is calculated as: $\sigma(h_i) = \frac{|E^+_{h_i}|}{|E^+|}$

In addition to the above, we also calculate for each hypothesis a measure, we call the *predictive value* of a hypothesis, which is based on the number of target selections the hypothesis infers for individuals with unknown target selections.

Definition 7
Given an ILP task $\langle B, E^+, E^-, M\rangle$ constructed using Def. 5, let $H = \{h_1, \cdots h_n\} \subseteq s(M)$ be an inductive solution to $E^+ \cup E^-$, w.r.t B. Let c_t be the predicate appearing in E. Let $|P|$ be the set of $c_t(id)$ atoms where c_t is Boolean, or $c_t(id,j)$ atoms where c_t is non-Boolean, with $id \in ID$, $j \in dom(c_t)$ and $\theta_{c_t}(id) = \{unknown\}$, entailed by $H \cup B$, such that $H \cup B \models p$ for every $p \in P$. Let $P_{h_i} \subseteq P$ be the set of $c_t(id)$, or $c_t(id,j)$, atoms entailed by $h_i \cup B$ where $h_i \in H$, such that $B \cup h_i \models p_{h_i}$, for every $p_{h_i} \in P_{h_i}$. Then the predictive ρ of h_i, denoted $\rho(h_i)$, is calculated as: $\rho(h_i) = \frac{|P_{h_i}|}{|P|}$.

Our algorithm uses the Answer Set solver `clingo` (Gebser et al. 2007) to find models of $B \cup h_i$ from which the relative significance and predictive values for each hypothesis are calculated. In summary, a logic program is constructed for each hypothesis $h_i \in H$ in conjunction with B. From this, the number of atoms representing target selections for each individual in the dataset in the answer set of the $B \cup h_i$ is computed. We distinguish between the number derived for those with known and unknown target selections by comparing the individuals' id's appearing in the answer set with those provided in the original dataset.

In our case, to learn characteristics and antecedent behaviour that are associated with specific target selections, B includes the clauses representing code assignments and relationships for the dataset of 111 lone-actor terrorists. The positive examples include facts about the observed target selection for each individual, e.g., $\{$*targetgroup*(*pg018*,*hvt*), *targetgroup*(*pg101*,*civilian*)$\}$ and the negatives includes facts about the groups that were not targeted, e.g., $\{$*targetgroup*(*pg018*,*civilian*), *targetgroup*(*pg101*,*hvt*)$\}$. The total number of positive examples is 95, and the total number of negative examples is 87.

The mode declarations M includes *modeh*(*targetgroup*(+*la_iden*, #*tg*)). It also comprises a total of 38 *modeb* declarations for codes such as *crimcon*, *verbfam*, *dryruns* and *mentalill* amongst others.

Having defined the ILP task, we use the ILP system XHAIL (Ray 2009), to compute the hypotheses. Table 1 shows an extract of the inductive solution. A full list of the mode declarations and hypotheses can be found at *www.doc.ic.ac.uk/∼da04/iclp15/*.

i	h_i			$\lvert e^+_{h_i}\rvert$	$\lvert p_{h_i}\rvert$
1	*targetgroup*(*I*, *civilian*)	←	*not crimcon*(*I*), *not history*(*I*), *not f2f*(*I*), *otherknowledge*(*I*), *not children*(*I*)).	4	2
2	*targetgroup*(*I*,*civilian*,)	←	*dryruns*(*I*), *not warning*(*I*), *not imprisoned*(*I*), *not training*(*I*), *mentalill*(*I*), *not virtualinteract*(*I*).	3	1
3	*targetgroup*(*I*,*civilian*)	←	*not imprisoned*(*I*), *not milexp*(*I*), *not livealone*(*I*), *not training*(*I*), *not virtuallearn*(*I*), *not otherknowledge*(*I*).	8	4
4	*targetgroup*(*I*,*civilian*)	←	*not milexp*(*I*), *verbfam*(*I*), *livealone*(*I*), *not f2f*(*I*).	4	4
5	*targetgroup*(*I*,*civilian*)	←	*uniexp*(*I*), *not warning*(*I*), *not milexp*(*I*), *not virtuallearn*(*I*).	12	1
...	...		...	...	...
52	*targetgroup*(*I*,*hvt*)	←	*not warning*(*I*), *mentalill*(*I*), *children*(*I*).	2	2
53	*targetgroup*(*I*,*hvt*)	←	*not imprisoned*(*I*), *virtuallearn*(*I*), *f2f*(*I*), *otherknowledge*(*I*), *recruit*(*I*) , *not children*().	1	0
54	*targetgroup*(*I*,*hvt*)	←	*not dryruns*(*I*), *not warning*(*I*), *not religcon*(*I*), *verbfam*(*I*), *not virtuallearn*(*I*), *not f2f*(*I*).	5	2
55	*targetgroup*(*I*,*hvt*)	←	*not univexp*(*I*), *not religcon*(*I*), *mentalill*(*I*), *not recruit*(*I*).	10	3
56	*targetgroup*(*I*,*hvt*)	←	*not* warning(*I*), *not imprisoned*(*I*), *crimcon*(*I*), *livealone*(*I*), *not training*(*I*), *not mentalill*(*I*), *not history*(*I*), *not recruit*(*I*).	1	0
			Unique total	95	29

Table 1: Hypothesis for target selection for the full sample

The solution shown above is not the minimal one. We have redefined the algorithm to terminate once it has found an optimal solution within a specified time bound. From the table above, we see that h_5 has a higher relative significance value than h_1 since $\sigma(h_5) = 0 \cdot 12 > \sigma(h_4) = 0 \cdot 04$, but a lower predictive value with $\rho(h_5) = 0 \cdot 03 < \rho(h_4) = 0 \cdot 13$. From a criminological perspective, the solution demonstrates that civilian targeting is associated with individuals with a history of mental illness (h_2) who engage in dry runs and have not been imprisoned, undergone training, amongst others. At the same time, we observe that mental illness alone cannot determine the target selection outcome as shown by (h_{55}) where mental illness exhibited with other characteristics, including no university experience or religious conversion prior to the attack, explains high-value target selections. Targeting high-value groups is also associated with criminal convictions and living alone which again may speak towards capability (in terms of both criminal ingenuity and having the space to develop a bigger plot). When this criminal ingenuity is not present, it may necessitate other behaviour like virtual learning, face to face interactions with co-ideologies and the attempt to recruit others (as witnessed in h_{52}).

3.3 Learning Characteristic and Behavioural Influences

The previous section is concerned with finding explanations using any subset of possible codes. In this section, we are interested in learning whether an observed behaviour can be explained with respect to particular characteristics and behaviour. We refer to this type of learning as *learning influences.*

A characteristic or behaviour is said to *influence the code* c_2 if every outcome of the observed behaviour represented by $c2$ can be explained in terms of the presence or absence of $c1$. In an ILP setting, the problem of learning influences is expressed in terms of finding inductive solutions in which every rule within that solution contains a body literal representing that characteristic or behaviour. For instance, in the case of exploring the influence of ideological orientation on target group selection, all hypotheses in an inductive solution must include a body literal corresponding to the individuals' ideological orientation.

Our algorithm for learning influences comprises four steps. First the dataset is split into subgroups, according to the value assigned to the code whose influence is being studied. Thus for a code c_f, we have $|dom(c_f)| \geq 2$ subgroups. Note that these subgroups are not required to be mutually exclusive. For ease of reference, we use the notation d_f to denote the subgroup containing data for individuals whose c_f value is d_f where the reference is obvious form the context. The second step involves applying the mapping in Def. 4 to each of the created subgroups separately. Then, a learning task $\langle B_{d_f}, E^+_{d_f}, E^-_{d_f}, M\rangle$ is defined for each subgroup d_f upon which the learning system XHAIL is executed. Once H_{d_f} is generated for each subgroup, the resulting inductive solutions undergo a post-processing procedure to generate the final set of hypotheses. This is done by applying a transformation function defined below to each rules in the inductive solutions.

Definition 8
Let Π be a normal logic program and b a literal. A transformation τ is defined such that $\Pi' = \tau(\Pi, b)$ and Π' is obtained from Π by adding a condition b to the body of every rule in Π.

Given the function τ, we have, in the case of a Boolean code c_f, the final solution $H = H'_{true} \cup H'_{false}$ where $H'_{true} = \tau(H_{true}, c_f(I))$, and $H'_{false} = \tau(H_{false}, not\ c_f(I))$. In the case of a non-Boolean code c_f, the final solution $H = \{H'_{d_{f_i}}\}$ where $H'_{d_{f_i}} =$

$\tau(H_{d_{f_i}}, c_f(I, d_{f_i}))$. Note that the correctness of solutions with respect to the union of the example sets is only guaranteed when the code values are independent.

In the case of learning the influence of ideological orientation on target selection, the dataset of 111 individuals is split into three subgroups based on whether the individual's ideological orientation is *rightwing*, *single_issue* or *religious*. In our example, the subgroups contain data for 43, 30, 38 individuals respectively. In our dataset, these groups are mutually exclusive since *singlevalue*(*ideology*)=*true*. The XHAIL system is then run using three independent learning tasks, one on each of the subgroups. The final solutions are shown in Table 2. The relative significance and predictive values are calculated with respect to each subgroup.

Religious Ideology

i	h_i			$\lvert e^+_{h_i}\rvert$	$\lvert p_{h_i}\rvert$
1	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,religious), not virtuallearn(I), not mentalill(I).*	10	2
2	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,religious), univexp(I), verbfam(I), not mentalill(I).*	7	0
3	*targetgroup(I,civilian)*	$\leftarrow$	*not univexp(I), f2f(I).*	3	0
4	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,religious),dryruns(I), otherknowledge(I).*	5	1
5	*targetgroup(I,hvt)*	$\leftarrow$	*ideo(I,religious),univexp(I), mentalill(I).*	6	3
6	*targetgroup(I,hvt)*	$\leftarrow$	*ideo(I,religious), not univexp(I), not warning(I), crimcon(I).*	9	2
7	*targetgroup(I,hvt)*	$\leftarrow$	*ideo(I,religious),not verbfam(I), virtuallearn(I).*	9	2
			Unique total	40	9

Single-issue Ideology

i	h_i			$\lvert e^+_{h_i}\rvert$	$\lvert p_{h_i}\rvert$
8	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,single_issue), not dryruns(I), not crimcon(I),not livealone(I).*	4	0
9	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,single_issue), crimcon(I), livealone(I).*	6	0
10	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,single_issue), f2f(I), not otherknowledge(I).*	11	0
11	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,single_issue),mentalill(I), not children(I).*	9	1
12	*targetgroup(I,hvt)*	$\leftarrow$	*ideo(I,single_issue), not livealone(I), not history(I), children(I).*	4	0
13	*targetgroup(I,hvt)*	$\leftarrow$	*ideo(I,single_issue),imprisoned(I), not mentalill(I), not recruit(I),not children(I).*	2	0
14	*targetgroup(I,hvt)*	$\leftarrow$	*ideo(I,single_issue),livealone(I), not mentalill(I), not history(I), not f2f(I).*	1	0
15	*targetgroup(I,hvt)*	$\leftarrow$	*ideo(I,single_issue), not crimcon(I), training(I).*	2	0
			Unique total	31	1

Rightwing Ideology

i	h_i			$\lvert e^+_{h_i}\rvert$	$\lvert p_{h_i}\rvert$
16	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,rightwing), not dryruns(I), livealone(I).*	14	4
17	*targetgroup(I,civilian)*	$\leftarrow$	*ideo(I,rightwing),verbfam(I), not mentalill(I), not children(I).*	7	1

18	*targetgroup(I,civilian)*	←	*ideo(I,rightwing),not training(I), not otherknowledge(I), not children(I).*	16	4
19	*targetgroup(I,civilian)*	←	*ideo(I,rightwing),not warning(I), mentalill(I), not recruit(I).*	10	2
20	*targetgroup(I,hvt)*	←	*ideo(I,rightwing), not livealone(I), children(I).*	5	3
21	*targetgroup(I,hvt)*	←	*ideo(I,rightwing), mentalill(I), not virtualinteract(I), otherknowledge(I).*	3	0
22	*targetgroup(I,hvt)*	←	*ideo(I,rightwing), not livealone(I), training(I), not f2f(I).*	1	0
23	*targetgroup(I,hvt)*	←	*ideo(I,rightwing), warning(I), mentalill(I).*	3	0
			Unique total	46	12

Table 2: Hypothesis for target selection for three ideological orientation groups.

The first part of Table 2 highlights a number of interesting facets related to a religious-inspired individual's choice. The presence (or lack thereof) of mental health problems helps shape target choice toward civilians or high-value targets respectively depending on whether the individual has university experience or not (see solution h_2 and solution h_5.) The lack of university experience (and perhaps the skills associated with overcoming complex tasks) can be mitigating for when targeting high-value targets by the presence of a criminal past and the nous that may develop through prior antecedent offending; see solution h_2. The second part in the table that refers to single-issue ideology indicates the need to disaggregate across ideological domains. Whereas the confluence of criminal histories and living alone appeared in Table 1 to suggest a close relationship with high-value targeting, the opposite is true for those individuals inspired by single issues (animal rights, environmentalists) and may be a direct reflection of different targeting norms within these movements (see solution h_9). In the absence of criminal histories, gaining training from a wider group appears to be a relevant substitute (see solution h_{15}). The last part of the table referring to rightwing ideologies confounds some expectations in the wider literature as it highlights the presence of mental health problems (solution h_{21}) in terms of attacking high-value targets compared to civilian targets (the latter of which are presumably easier to plan).

4 Discussion

In our approach, we focused on learning rules that capture associations between behaviour and influence of specific behaviour when explaining terrorists' target selection. The performance of the learning algorithm used depended on the size of M amongst other factors such as the size of the examples. Experiments showed that it decreased when a larger M was considered. The performance of the algorithm for same M size was higher in the case of learning influences than it was when learning associations, as expected given the examples size was smaller. Furthermore, we found that we were able to find a more optimal solution for learning influences than we did for associations for the fixed time-frame we gave (which was set to

360 minutes) for the same observation (56 rules in the case of learning associations compared to 23 when learning influences). The choice of codes to use in M in the presented work was influenced by variables that are commonly investigated in existing literature. Furthermore, the optimization was driven by the total number of literals appearing in an inductive solution in some cases resulting in more hypotheses with small potential significance values as opposed to fewer hypotheses with higher potential significance. The learning association approach is similar in essence to task discovery in data-mining, e.g.,(Dehaspe and Toivonen 1999), the aim here is to provide a reasoning platform capable of handling default negation which better suits the incremental hypothesis generation and refinement nature of the problem domain, and allows the integration of domain knowledge both in the background and in the heuristics defined over the search space.

To evaluate our approach, we compared our results against those produced using standard methods deployed in terrorism studies, rather than performed cross validation over our small sample set. In particular, we conducted a Smallest Space Analysis (SSA), shown in Fig. 1, of the antecedent behaviour and their relationship to one another for the full dataset of 111 individuals. Such analyses focus upon variable co-occurrence. Prominent examples include (Canter and Heritage 1990) work on serial rapists, and (Canter 2004) work on serial murder. The lone-actor terrorist typologies presented below utilises this specific method. It provides geometric representations of the level of association between variables. In other words, the Multi Dimensional Scaling outputs represent a matrix wherein variables that regularly co-occur are plotted closer together in a Euclidean space. The utility of such a representation is that the variable configuration is based upon variables' relationships with each other rather than their relationships with pre-determined dimensions (Davis 2009). SSA is based upon the assumption that the underlying structure of complex systems is most readily appreciated if the relationship between each and every other variable is examined, but that such examination is much clearer if the relationships are represented visually not only in terms of numbers' (Canter 2004). The Jaccard co-efficient (which represents the level of association between two variables) was calculated for each pair-wise set of variables. The closer two variables appear within the matrix, the higher their co-occurrence across observations. For example, virtual learning (VirtualLearn) and virtual interaction (VirtualInteract) are extremely close and therefore occur very regularly together.

The results visually illustrate some of the key findings produced in Table 1. The SSA output also helps demonstrate which of the "not" behaviour rarely co-occur with the present behaviour and which are specific to that combination of factors. As per h_1, other knowledge is situated very far from criminal convictions, history of violence, children indicated that this particular behaviour rarely co-occurs with these other factors. However, the SSA output demonstrates a relatively close relationship between other knowledge and face-to-face interactions. h_1 however shows these rarely co-occur when these other factors are also absent. The SSA output also helps illustrate the degree to which the "not" behaviour co-occur. Returning to h_1, children rarely co-occur with a history of violence. The types of hypotheses that the SSA struggles with are those that are purely made up of "not" occurrences because what underpins the SSA is the co-occurrence of two variables. To illustrate, h_3 is very difficult to comprehend using the SSA.

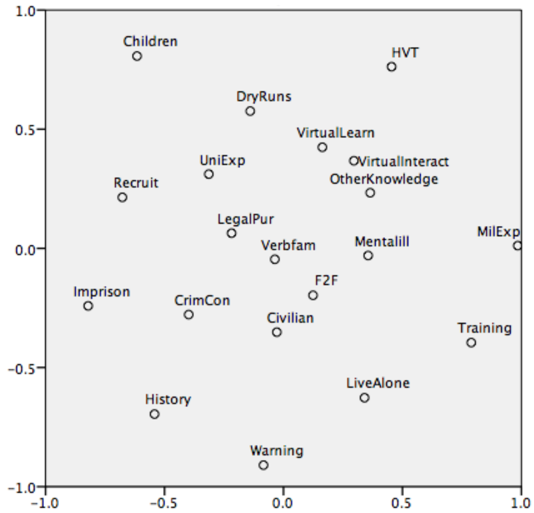

Fig. 1. Smallest Space Analysis results for 111 lone actor terrorists.

5 Conclusion and Future Work

The aim of this paper is to examine the applicability of ILP for the purpose of generating relevant hypothesis about terrorists' behaviour. The findings reported in the previous sections collectively show how ILP not only provides the ability to derive new insights but it is also clearly beneficial in outlining the explanatory power of rules. Whilst the SSA approach outlines the many diffuse relationships between a sizeable number of variables, it is very difficult to focus on the most relevant ones. The clusters that tend to emerge through identification by research teams tend to be quite subjective - more art than science - and are therefore subject to potential bias. The SSAs also tend to drag commonly occurring variables into the centre of the model whilst non-common variables are pushed to the extremes. The ILP approach, we believe, has the power to delineate which relationships are highly relevant and more immediately usable.

Our ongoing and future work includes distinguishing between causal and non-causal factors when generating solutions, prioritizing hypotheses with causal explanations and higher relative significance values, and applications to other criminological problems such as serial crimes. We plan to conduct further investigation into prioritizing and optimizing the selection of the body literals when constructing hypotheses, one possibility is by using the results from SSA or information gain theory . We also intended to investigate the use of probabilistic learning and methods capable of handling noise and uncertainty in this setting.

References

CANTER, D. 2004. Offender profiling and investigative psychology. *Journal of Investigative Psychology and Offender Profiling 1,* 1, 1–15.

CANTER, D. AND HERITAGE, R. 1990. A multivariate model of sexual offence behaviour: developments in offender profiling. *Journal of Forensic Psychiatry 1*, 185–212.

DAVIS, M. 2009. In defence of multidimensional scaling for the analysis of sexual offence behaviour: cautionary notes regarding analysis and interpretation. *Psychology, Crime & Law 15,* 6, 507–515.

DEHASPE, L. AND TOIVONEN, H. 1999. Discovery of frequent datalog patterns. *Data Mining and Knowledge Discovery 3,* 1, 7–36.

FARRINGTON, D. P. 2000. Explaining and preventing crime: The globalisation of knowledge — the ASC 1999 presidential address. *Criminology 38,* 1, 1–24.

GEBSER, M., KAUFMANN, B., NEUMANN, A., AND SCHAUB, T. 2007. *clasp* : A conflict-driven answer set solver. In *Proceedings of 9th International Conference on Logic Programming and Nonmonotonic Reasoning*. 260–265.

GELFOND, M. AND LIFSCHITZ, V. 1988. The stable model semantics for logic programming. In *Proceedings of the 5th International Conference on Logic Programming*. The MIT Press, 1070–1080.

GILL, P. AND CORNER, E. 2013. Disaggregating terrorist offenders: Implications for research and practice. *Criminology & Public Policy 12,* 1, 93–101.

GILL, P., HORGAN, J., AND DECKERT, P. 2014. Bombing alone: Tracing the motivations and antecedent behaviors of lone-actor terrorists. *Journal of Forensic Sciences 59,* 2, 425–435.

HORGAN, J. 2014. *The Psychology of Terrorism*, 2 ed. Routledge, Oxford, UK.

LAFREE, G. 2013. Lone-offender terrorists. *Criminology & Public Policy 12,* 1, 59–62.

MUGGLETON, S. 1991. Inductive logic programming. *New Generation Computing 8,* 4 (Feb.), 295–318.

RAY, O. 2009. Nonmonotonic abductive inductive learning. *Journal of Applied Logic 7,* 3, 329–340.

RAY, O., BRODA, K., AND RUSSO, A. 2004. A hybrid abductive inductive proof procedure. *Logic Journal of the IGPL 12,* 5, 371–397.

SAKAMA, C. AND INOUE, K. 2009. Brave induction: a logical framework for learning fromincomplete information. *Machine Learning 76,* 1, 3–35.

WIKSTROM, P. 2011. Does everything matter? Addressing the problem of causation and explanation in the study of crime. In *When Crime Appears: The Role of Emergence*, J. McGloin, C. Sullivan, and L. Kennedy, Eds. Routledge.