

Aberystwyth University

Evidence directed generation of plausible crime scenarios with identity resolution

Fu, Xin; Shen, Qiang; Boongoen, Tossapon

Published in:

Applied Artificial Intelligence

DOI:

[10.1080/08839511003715154](https://doi.org/10.1080/08839511003715154)

Publication date:

2010

Citation for published version (APA):

Fu, X., Shen, Q., & Boongoen, T. (2010). Evidence directed generation of plausible crime scenarios with identity resolution. *Applied Artificial Intelligence*, 24(4), 253-276. <https://doi.org/10.1080/08839511003715154>

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400

email: is@aber.ac.uk

This article was downloaded by: [Fu, xin]

On: 12 April 2010

Access details: Access Details: [subscription number 921256321]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Applied Artificial Intelligence

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713191765>

EVIDENCE DIRECTED GENERATION OF PLAUSIBLE CRIME SCENARIOS WITH IDENTITY RESOLUTION

Xin Fu ^a; Tossapon Boongoen ^a; Qiang Shen ^a

^a Department of Computer Science, Aberystwyth University, Penglais Campus, Aberystwyth, United Kingdom

Online publication date: 12 April 2010

To cite this Article Fu, Xin , Boongoen, Tossapon and Shen, Qiang (2010) 'EVIDENCE DIRECTED GENERATION OF PLAUSIBLE CRIME SCENARIOS WITH IDENTITY RESOLUTION', Applied Artificial Intelligence, 24: 4, 253 – 276

To link to this Article: DOI: 10.1080/08839511003715154

URL: <http://dx.doi.org/10.1080/08839511003715154>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

EVIDENCE DIRECTED GENERATION OF PLAUSIBLE CRIME SCENARIOS WITH IDENTITY RESOLUTION

Xin Fu, Tossapon Boongoen, and Qiang Shen

*Department of Computer Science, Aberystwyth University, Penglais Campus,
Aberystwyth, United Kingdom*

□ *Given a set of collected evidence and a predefined knowledge base, some existing knowledge-based approaches have the capability of synthesizing plausible crime scenarios under restrictive conditions. However, significant challenges arise for problems where the degree of precision of available intelligence data can vary greatly, often involving vague and uncertain information. Also, the issue of identity disambiguation gives rise to another crucial barrier in crime investigation. That is, the generated crime scenarios may often refer to unknown referents (such as a person or certain objects), whereas these seemingly unrelated referents may actually be relevant to the common revealed. Inspired by such observation, this article presents a fuzzy compositional modeler to represent, reason, and propagate inexact information to support automated generation of crime scenarios. Further, the article offers a link-based approach to identifying potential duplicated referents within the generated scenarios. The applicability of this work is illustrated by means of an example for discovering unforeseen crime scenarios.*

INTRODUCTION

To resolve the puzzle of a given crime from a set of available evidence, investigators and forensic analysts aim to disclose the scenario that has actually taken place and to determine efficient strategies to proceed with the investigation (Shen et al. 2006). Hence, the effectiveness of this analytical process depends critically on investigators' ability to articulate plausible scenarios and to identify course of actions needed for differentiating them.

Although the bottleneck of accomplishing such a task is the fact that humans are relatively inefficient at hypothetical reasoning, a

This work is partly sponsored by the UK EPSRC grant no. EP/D057086. We are grateful to the members of the project team for their contribution and take full responsibility for the views expressed in this article.

Address correspondence to Xin Fu, Department of Computer Science, Aberystwyth University, Penglais Campus, Aberystwyth SY23 3DB, UK. E-mail: xxf06@aber.ac.uk

decision-support system may assist investigators in generating plausible scenarios and analyzing them objectively. To this extent, compositional modeling (CM) (Keppens and Shen 2001) has been successfully used to synthesize plausible crime scenarios from collected evidence (Keppens et al. 2005; Shen et al. 2006). This intelligent CM-based approach is brought forward to underpin the significant advantage of its ability to automatically construct many variations of an underlying crime scenario from a relatively small knowledge base. However, in such existing work it is assumed that the basic knowledge elements, termed scenario fragments (SF) as collected within the knowledge base, can all be expressed by precise and crisp information. Particularly, a set of probability distributions, possessed by each SF for representing the likelihood of its associated possible outcomes, are specified in numerical forms. In spite of this, especially in the case of crime detection and prevention, the degree of precision of available evidence and intelligence data can vary greatly, subject to different perception, judgement, and individuality of people. As such, an assessment of likelihood typically reflects the expertise and knowledge of experienced investigators is normally available in linguistic terms instead (Halliwell et al. 2003; Halliwell and Shen 2009). The use of seemingly precise numeric descriptors suffers from an inadequate degree of accuracy.

Another crucial barrier for any investigation decision-support system is *identity disambiguation problem* that is commonly encountered in intelligence data analysis (Badia and Kantardzic 2005; Pantel 2006; Wang et al. 2005). In particular, a modest resolution used within the previous crime investigation systems (Keppens et al. 2005; Shen et al. 2006) is to assume that the identities of any involving instances (e.g., person, object, place, and organization) are globally unambiguous. However, this simple ignorance of possible identity aliases drastically reduces the degree of this approach's flexibility and actual utilization within the real world. Yet, another significant problem occurs whenever the scenario composition process has to refer to unknown identities. For instance, given the evidence of a man having been shot by a 9-mm gun, the abductive reasoner (Shen et al. 2006) can hypothesize the existence of a murderer owning such a weapon but requiring the possible matching of this unknown gun with weapon identities instantiated within the scenario space. Henceforth, complementing such abductive reasoning with the mechanism of identity resolution would enhance not only the quality and coverage of scenario acquired by associating sparse SFs through shared identities of involving participants but also the effectiveness of system management regarding duplicated fragments.

Recognizing the aforementioned shortcomings, we extend the initial model of *fuzzy CM* (Fu et al. 2007) in which the theory of fuzzy sets is applied to the conventional CM approach to provide a mechanism

for representing, reasoning, and propagating vague and uncertain information. In addition, a link-based approach to identity resolution, which has proven effective for disclosing aliases in intelligence data (Boongoen and Shen 2008; Boongoen et al.), is exploited to identify duplicated identities within a collection of generated scenarios. This novel approach is not only able to refine the quality of scenarios but also to optimize the resource utilization.

The rest of this article is organized as follows. We next introduce knowledge representation of vague and uncertain information within CM by the use of fuzzy sets and then present an outline of the proposed approach. Then, we show how the plausible scenarios can be synthesized from available evidence and knowledge base under uncertainty. Next we describe how such generated scenarios can be refined to handle the identity problem by using link-based analysis. This is followed by an illustrative example showing the capability of the proposed approach to provide support in crime investigation. Finally, we conclude this article and point out future research.

REPRESENTING EVIDENCE AND CRIME SCENARIOS UNDER UNCERTAINTY

Knowledge Representation in CM

The conventional CM approach has been successfully applied to generate descriptions of crime scenarios from evidence (Keppens et al. 2005). In CM the knowledge base consists of a number of generic SFs, which represent generic relationships between domain objects and their states for certain types of partial scenario. An SF has two parts that encode domain knowledge: i) the relations between domain elements, which are often represented in a form that is similar to the style of conventional production rules but involving much more general contents, and ii) a set of rule instances that represent how likely it is that the corresponding relationships hold.

More formally, an SF is represented as follows:

If $\{P^s\}$
Assuming $\{P^a\}$
Then $\{p^t\}$
Distribution $p^t\{$
 \dots
 $v_1^s, \dots, v_m^s, v_1^a, \dots, v_k^a \rightarrow v^t : q$
 $\dots\}$

where

- $\mathbf{P}^s = \{p_1^s, \dots, p_m^s\}$ is a set of antecedent predicates. The **If** statement describes the required conditions for a partial scenario to become applicable.
- $\mathbf{P}^a = \{p_1^a, \dots, p_k^a\}$ is a set of assumption predicates, referring to those pieces of information which are unknown or cannot be inferred from other scenario fragments, but they may be presumed to be known for the sake of performing hypothetical reasoning.
- p^t is the consequent predicate, which describes the consequent when the conditions and presumed assumptions hold. It may represent a piece of new knowledge or relations which are derived from the hypothetical reasoning.
- In the **Distribution** statement, the left-hand side of the “implication” sign in each rule instance is a combination of predicate-value pairs, involving antecedent and assumption predicates and their values, and the right-hand side indicates the likelihood of each alternative outcome if the fragment is instantiated.

Given a collection of such SFs and some collected pieces of evidence, CM applies an abductive reasoner to create plausible scenarios (see Shen et al. 2006 for details). However, for many problem domains, especially for crime detection and prevention, different kinds of inexact information may be involved which conventional CM cannot always handle, including

- **Vagueness:** It occurs when the boundary of a piece of information cannot be determined precisely (e.g., Bob is *tall*).
- **Uncertainty:** It captures the reliability or confidential weight of a given piece of information, usually represented by numerical values (e.g., If X is a bird, then it can fly, with certainty degree 0.95).
- **Both vagueness and uncertainty:** Often, these two different kinds of inexactness may coexist (e.g., The amount of collected fiber is *a lot*, with certainty degree 0.7).
- **Both vagueness and uncertainty with the uncertainty also described in fuzzy terms:** Instead of using numerical values, the uncertainty is represented as fuzzy numbers or linguistic terms (e.g., The amount of collected fiber is *a lot*, with certainty degree *very likely*).

The following subsections focus on the creation of a structured knowledge representation scheme, which is capable of storing and managing both vague and uncertain information. It involves two conceptually distinct aspects: i) fuzzification of SFs, including fuzzy parameters and fuzzy constraints, and ii) definition of fuzzy complex numbers, which facilitates the propagation of vagueness and uncertainty within the process of model composition.

Fuzzification of SFs

Fuzzy Parameters

Within the present work, vague and ill-defined concepts and quantities are defined as fuzzy variables whose values are fuzzy sets. For example, the police discovered the dead body of Smith in his bedroom. Bob, who is the next-door neighbor, witnessed somebody going into Smith's house; however, it is difficult for Bob to state an accurate height for that person (e.g., 180 cm). Intuitively, he might just describe the height of the person as *short*, *average*, or *tall*. In this case, *height* can be modeled as a fuzzy variable and can be described by using fuzzy sets *short*, *average*, or *tall*.

In implementation, fuzzy variables are indicated by means of the keyword **fuzzyvariable**. Defining such a variable involves specifying the following fields:

- **Name:** A constant that uniquely identifies the fuzzy variable.
- **Universe of discourse:** The domain of the fuzzy variable.
- **Unit:** The variable's physical dimension. If a fuzzy variable has no unit, a default value of *none* is set.
- **Cardinality of partition:** The number of fuzzy sets that jointly partition the universe of discourse. It is represented by a positive integer.
- **Quantity space:** The membership functions of the fuzzy sets that jointly cover the partitioned domain.
- **Name of fuzzy sets:** The symbolic label of each fuzzy set in the defined quantity space.
- **Unifiability:** The declaration of a unifiable property of the variable, specified by a predicate.

The following example defines a fuzzy variable named *Chance*. The defined quantity space is shown in Figure 1.

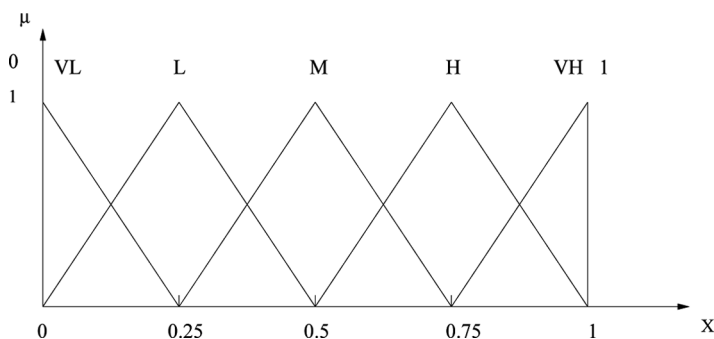


FIGURE 1 The fuzzy quantity space of *Chance*.

Define fuzzyvariable{

Name : Chance

Universe of discourse : $[0, 1]$

Unit : none

Cardinality of partition : $n = 7$

Quantity space :

$$fs_1 = [0, 0, 0]$$

$$fs_2 = \left[0, 0, \frac{1}{n-3} \right]$$

...

$$fs_i = \left[\frac{i-3}{n-3}, \frac{i-2}{n-3}, \frac{i-1}{n-3} \right]$$

...

$$fs_{n-1} = \left[\frac{n-4}{n-3}, 1, 1 \right]$$

$$fs_n = [1, 1, 1]$$

Name of fuzzy sets : $\{0, VL, L, M, H, VH, 1\}$

Unifiability : Chance(X)

}

Fuzzy Constraints

In CM, knowledge is normally expressed as constraints or relations that must be obeyed by certain variables. Particularly, constraints used in the existing work require numerical values to quantify the probability of a consequence's occurrence. However, such subjective assessments are often the product of barely articulate intuitions. In fact, the seemingly numerically precise expressions may cause loss of efficiency, accuracy, and transparency (Cooman 2005; Halliwell et al. 2003; Halliwell and Shen 2009). Often, an expert may be unwilling or simply unable to suggest a numerical probability.

The work developed here models the vagueness of the probability distribution in terms of subjective certainty degree. Rather than using numerical value as in the literature, a fuzzy variable called *Chance*, which takes linguistic values from $[0, 1]$, is introduced to capture subjective certainty degrees.

The following sample SF illustrates the concepts and applicability of fuzzy parameters and fuzzy constraints:

If {height(S), height(V)}

Assuming {attempted to kill(S, V)}

Then {difficult level(overpower(S, V))}
Distribution difficult level(overpower(S, V)){
 r1: tall, short, true \rightarrow easy: VH
 r2: short, tall, true \rightarrow easy: VL}.

It describes a general relation between the heights of two people involved in a fight and the difficulty level for one to overpower the other. Here, the *height* is modeled as a fuzzy variable that takes values from a predefined quantity space of $Q = \{very\ short, short, average, tall, very\ tall\}$, and *difficulty level* is another fuzzy variable with possible values being *easy*, *average*, and *difficult*. In particular, this fragment covers two rule instances indicating that if suspect S is *tall* and victim V is *short* and S indeed attempted to kill V , then S stands a *very high* chance to overpower V easily. Conversely, if S is shorter than V and he indeed attempted to kill V , then there is only a *very low* chance for S to overpower V easily.

Note that it is not required that every possible combination of antecedent and assumption values has to be assigned a certainty degree because the number of combinations increases exponentially with the number of variables. Knowledge embedded in any reasoning system is always incomplete, and it is unlikely to obtain all such details but the most significant components. As with any practical knowledge-based approach, the default certainty degree for those unassigned combinations is set to 0.

Fuzzy Complex Numbers

Because generic SFs may involve both vague and uncertain information, when dynamically composing the potentially relevant ones into plausible scenario descriptions, such inexact information needs to be combined and propagated in the emerging scenarios. To achieve this, an innovative notion of fuzzy complex numbers (FCN), which extends real complex numbers, has recently been proposed (Fu and Shen 2009) with the significant capability of representing two-dimensional uncertainties simultaneously. For completeness, a brief account of this novel concept is given below in the context of CM.

Definition 1. Inherit from the real complex numbers, an FCN, \tilde{z} , is defined in the form of:

$$\tilde{z} = \tilde{a} + i\tilde{b}, \quad (1)$$

where both \tilde{a} and \tilde{b} are fuzzy numbers with membership functions $\mu_{\tilde{a}}(x)$ and $\mu_{\tilde{b}}(x)$. It is obvious that \tilde{a} is the real part of \tilde{z} , whereas \tilde{b} represents the imaginary part, i.e., $Re(\tilde{z}) = \tilde{a}$ and $Im(\tilde{z}) = \tilde{b}$.

In this work, to support CM both the real part and the imaginary part of an FCN are assigned with their embedded semantic meaning, and they will be propagated, respectively, in the model composition process. Specifically, the real part represents the certainty degree of a certain piece of information, whereas the imaginary part represents the fuzzy matching degree of a given evidence or derived piece of information using a particular rule instance of a certain SF. As such, FCNs offer an effective and efficient common scheme to represent the four types of inexact information as previously identified such that

- **Vagueness:** In CM, the imaginary part only exists when fuzzy matching is performed between two pieces of information. In addition, given a piece of evidence, if there is no specific certainty degree assigned to it, then it is assumed to be certain by default. In this case the real part of the FCN in representing this piece of evidence is equal to 1. For example, given a fact, f_1 : Bob is *tall*, this single piece of vague information is represented as $\tilde{z} = 1$. However, when a piece of evidence, e_1 : victim is *tall*, is collected, f_1 matches with e_1 with a fuzzy matching degree \tilde{b} , such that if f_1 is activated, the FCN attached to f_1 is written as $\tilde{z} = 1 + i\tilde{b}$.
- **Uncertainty:** If a piece of information (e.g., a rule) only involves uncertainty, then the corresponding uncertainty measure can be represented as $\tilde{z} = \tilde{a}$, which can either be a real number or a fuzzy number.
- **Both vagueness and uncertainty coexist:** The third and fourth types of inexactness can both be represented in the following generic form: $\tilde{z} = \tilde{a} + i\tilde{b}$. The only difference is that \tilde{a} is a real number in the third type and is a fuzzy number in the fourth.

Note that the basic operations on FCNs form a straightforward extension of those on real complex numbers, and they are omitted here. Interested readers can refer to (Fu and Shen 2009) for further details. However, it is worthy to point out that the modulus of FCNs is introduced to constrain the scenario descriptions; it will be used as a global measure to evaluate the overall confidence level of a piece of information. As such, the modulus of an FCN \tilde{z} is defined as

$$|\tilde{z}| = \sqrt{\tilde{a}^2 + \tilde{b}^2}, \quad (2)$$

where $|\tilde{z}|$ is a newly derived fuzzy number with the following membership function:

$$\mu_{|\tilde{z}|}(y) = \bigvee_{y=\sqrt{x_1^2+x_2^2}} (\mu_{\tilde{a}}(x_1) \wedge \mu_{\tilde{b}}(x_2)). \quad (3)$$

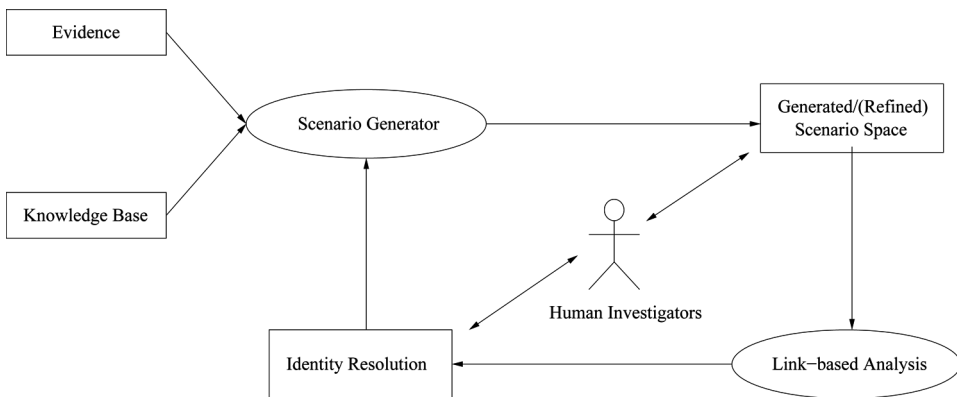


FIGURE 2 Outline of the approach.

OUTLINE OF THE APPROACH

The outline of the proposed approach is shown in Figure 2. A new or ongoing investigation results in an initial set of evidence. Then, the *Scenario Generator* component generates a collection of plausible scenarios, called scenario space, by exploiting a given knowledge base. The scenario space describes how plausible states and events may be related to have caused the given evidence. Having obtained scenarios, the *Link-based Analysis* component is used to handle the identity problem. Consequently, a list of duplicate candidates is fed back to the *Scenario Generator* to produce an updated, more refined, scenario space.

GENERATION OF PLAUSIBLE SCENARIOS

Given a set of collected evidence and a predefined knowledge base, plausible scenarios can be generated by a joint exploitation of two conventional inference techniques, named *backward chaining* and *forward chaining*. Note that because the precision degree of the underlying information can vary substantially, the collected evidence and the knowledge base in general cannot be matched precisely, and a partial matching may suffice. A fuzzy matching method is developed to perform such partial matching and hence SF instantiation (with respect to the given evidence). In addition, when composing an instantiated SF with an emerging plausible scenario, the attached certainty degree and fuzzy matching degree (which are concisely represented in terms of FCNs) need to be propagated from individual SFs to their related ones. Hence, an algorithm for propagating FCNs within the model composition process is also proposed herein.

Initialisation

Collected evidence and facts are used to form the initial elements contained within the emerging scenarios space (which is empty before this initialisation step). To allow instantiation of a fuzzy SF in accordance with a given piece of evidence, the system requires matching specific data items with broader and relatively subjective information. Such fuzzy matching is accomplished by the following procedure:

1. Find the so-called candidate SFs that involve any of the variables used to describe the collected evidence.
2. Identify the degree of match between evidence and each candidate SF. For simplicity, the match degree is obtained by calculating the maximum membership value over the overlapping area between the relevant fuzzy sets.
3. Return the matched SFs with the largest FCN modulus for instantiation.

Propagation of FCNs

After the initial instantiation, each added node in the emerging space has an FCN attached. Note that each rule instance in any generic SF also has an FCN attached that is called \tilde{z}_r , which indicates the certainty degree of its subjective causal propositions. In the generation of plausible scenarios, the attached FCNs are propagated to their related nodes via the FCN of the activated rule instance (\tilde{z}_r).

For simplicity, the following rule instance in an SF:

$$\text{IF } p_1^s \text{ is } v_1^s \oplus \cdots \oplus p_m^s \text{ is } v_m^s, \quad \text{ASSUMING } p_1^a \text{ is } v_1^a \oplus \cdots \oplus p_k^a \text{ is } v_k^a, \\ \text{THEN } p^t \text{ is } v^t(\tilde{z}_r)$$

is rewritten as

$$\text{IF } p_1 \oplus \cdots \oplus p_n, \quad \text{THEN } c(\tilde{z}_r),$$

where p_1, \dots, p_n , $n = m + k$ are the antecedent propositions, c is the consequent proposition, and \tilde{z}_r is the FCN attached to the rule. Note that the assumption propositions are treated as part of the antecedent due to their logical equivalence. Also, \oplus in the antecedent may be interpreted as either a conjunctive operator or a disjunctive operator.

Definition 2. If \oplus is a conjunctive operator, the aggregated FCN of the antecedent is

$$\tilde{z}_{\text{antecedent}} = \min(\text{Re}(\tilde{z}_{p_1}), \dots, \text{Re}(\tilde{z}_{p_n})) + i \min(\text{Im}(\tilde{z}_{p_1}), \dots, \text{Im}(\tilde{z}_{p_n})). \quad (4)$$

Definition 3. If \oplus is a disjunctive operator, the aggregated FCN of the antecedent is

$$\tilde{z}_{\text{antecedent}} = \max(\text{Re}(\tilde{z}_{p_1}), \dots, \text{Re}(\tilde{z}_{p_n})) + i \max(\text{Im}(\tilde{z}_{p_1}), \dots, \text{Im}(\tilde{z}_{p_n})). \quad (5)$$

The propagation process is then a combination process of the FCNs given to the variables (antecedents or consequent) with the FCN attached to the activated rule instances (\tilde{z}_r). Because the rule instance in an SF only has the certainty degree attached, i.e., $\tilde{z}_r = \text{Re}(\tilde{z}_r)$, the newly derived FCN attached to the consequent is computed using the following method:

$$\begin{aligned} \tilde{z}_{\text{new}} &= \tilde{z}_{\text{antecedent}} \times \tilde{z}_r \\ &= \text{Re}(\tilde{z}_{\text{antecedent}}) \times \tilde{z}_r + i \text{Im}(\tilde{z}_{\text{antecedent}}) \times \tilde{z}_r, \end{aligned} \quad (6)$$

where $\text{Re}(\tilde{z}_{\text{antecedent}})$, $\text{Im}(\tilde{z}_{\text{antecedent}})$, and \tilde{z}_r are represented by different fuzzy numbers.

Backward Chaining

This involves the *abduction* of domain variables and their states, which might have led to the available evidence. Plausible causes are created by instantiating the conditions and assumptions of the SFs, whose consequent match the given evidence with the largest FCN modulus. After that, the newly created instances of all plausible causes are recursively used in the same manner as with the original piece of evidence, instantiating all relevant fragments and adding the resulting instantiated variables, termed nodes to the emerging scenario space.

Due to the lack of inverse operators over fuzzy numbers, in backward chaining, given \tilde{z}_c and \tilde{z}_r , no general method exists to derive the exact value of $\tilde{z}_{\text{antecedent}}$ in a closed form. Hence, specific fuzzy quantity spaces are built to approximately represent the possible values of the real and imagery parts of $\tilde{z}_{\text{antecedent}}$, assuming that they can only take values from such predefined quantity spaces. For computational simplicity, in this work it is further assumed that only fuzzy values from the same quantity space, $Q_{FN} = \{V_{FN_1}, \dots, V_{FN_j}, \dots, V_{FN_n}\}$, are possibly taken by \tilde{z}_c and \tilde{z}_r .

In backward chaining, \tilde{z}_c and \tilde{z}_r are given. Algorithm 1 is constructed by ensuring the generality that each element in Q_{FN} could be the possible value of $\tilde{z}_{\text{antecedent}}$. Thus, all such values are checked using Eq. (6) as the constraint. The value that best satisfies this constraint when given \tilde{z}_c is selected to represent $\tilde{z}_{\text{antecedent}}$. Note that, the $\text{Re}(\tilde{z}_c)$ and $\text{Im}(\tilde{z}_c)$ are checked against the constraint separately.

Forward Chaining

All plausible causes of the collected evidence that the knowledge base entails may be introduced to the emerging scenario space during the

Algorithm 1 Backward propagation (\tilde{z}_c, \tilde{z}_r)

```

1:  $Re(\tilde{z}_{\text{antecedent}}) \leftarrow \emptyset$ 
2:  $Im(\tilde{z}_{\text{antecedent}}) \leftarrow \emptyset$ 
3:  $S_{Re} = 0$ 
4:  $S_{Im} = 0$ 
5: for  $j = 1 : n$  do
6:   if  $S(Re(\tilde{z}_c), V_{FN_j} \times \tilde{z}_r) > S_{Re}$  then
7:      $S_{Re} = S(Re(\tilde{z}_c), V_{FN_j} \times \tilde{z}_r)$ 
8:      $Re(\tilde{z}_{\text{antecedent}}) = V_{FN_j}$ 
9:   end if
10:  if  $S(Im(\tilde{z}_c), V_{FN_j} \times \tilde{z}_r) > S_{Im}$  then
11:     $S_{Im} = S(Im(\tilde{z}_c), V_{FN_j} \times \tilde{z}_r)$ 
12:     $Im(\tilde{z}_{\text{antecedent}}) = V_{FN_j}$ 
13:  end if
14: end for
15: return  $\tilde{z}_{\text{antecedent}}$ 

```

backward chaining phase. Here, the forward chaining phase is responsible for extending the scenario space by adding all plausible consequences of the fragments whose conditions and assumptions match the instances created in the last phase. This produces further pieces of information that have not yet been identified otherwise but may be used to improve the plausible scenario description.

This procedure applies logical *deduction* to all the SFs in the knowledge base, whose conditions match the existing nodes in the emerging scenario space. Therefore the corresponding FCN propagation algorithm is straightforward, as described in Algorithm 2. In particular, $\tilde{z}_{\text{antecedent}}$ and \tilde{z}_r are known, the \tilde{z}_{new} is computed by applying Eq. (6), where \tilde{z}_{new} is the calculated FCN of the consequent variable. Given the fired Q_{FN} in this work, once \tilde{z}_{new} is obtained it is used to match with the elements in Q_{FN} , the element receives the highest fuzzy matching degree with \tilde{z}_{new} is returned to represent \tilde{z}_c .

FCN Updating

In the emerging scenario space, if an existing node resulted from instantiating another SF such that two instantiated SFs share a common variable, then the existing FCN of this node needs to be updated. Suppose that given the following two FCNs:

$$\begin{aligned}
 \tilde{z}_A &= V_{FN_i} + iV_{FN_j}, \\
 \tilde{z}'_A &= V_{FN_m} + iV_{FN_k}.
 \end{aligned} \tag{7}$$

Algorithm 2 Forward propagation ($\tilde{z}_{p_1}, \dots, \tilde{z}_{p_m}, \tilde{z}_r$)

```

1: Calculate  $\tilde{z}_{\text{antecedent}}$ 
2:  $\tilde{z}_{\text{new}} = \tilde{z}_{\text{antecedent}} \times \tilde{z}_r$ 
3:  $S_{Re} = 0$ 
4:  $S_{Im} = 0$ 
5: for  $j = 1 : n$  do
6:   if  $S(V_{FN_j}, Re(\tilde{z}_{\text{new}})) > S_{Re}$  then
7:      $S_{Re} = S(V_{FN_j}, Re(\tilde{z}_{\text{new}}))$ 
8:      $Re(\tilde{z}_c) = V_{FN_j}$ 
9:   end if
10:  if  $S(V_{FN_j}, Im(\tilde{z}_{\text{new}})) > S_{Im}$  then
11:     $S_{Im} = S(V_{FN_j}, Im(\tilde{z}_{\text{new}}))$ 
12:     $Im(\tilde{z}_c) = V_{FN_j}$ 
13:  end if
14: end for
15: return  $\tilde{z}_c$ 

```

As the modulus of an FCN is used to evaluate the information content, if $|\tilde{z}_A|$ and $|\tilde{z}'_A|$ are the same (while $\tilde{z}_A \neq \tilde{z}'_A$), then both \tilde{z}_A and \tilde{z}'_A are kept as possible FCNs. However, if $|\tilde{z}_A| \neq |\tilde{z}'_A|$, then the newly updated FCN of variable X_1 taking value A is obtained by

$$\tilde{z}''_A = \min(V_{FN_i}, V_{FN_m}) + i \min(V_{FN_j}, V_{FN_k}). \quad (8)$$

This is an intuitive appeal because, if $\tilde{z}''_A = \min(V_{FN_i}, V_{FN_m}) + i \min(V_{FN_j}, V_{FN_k})$ is supported by the existing inference, then $\tilde{z}''_A = \max(V_{FN_i}, V_{FN_m}) + i \max(V_{FN_j}, V_{FN_k})$ is also supported.

Removal of Spurious Nodes

In the previous phases of scenario space generation, spurious nodes may have been added to the emerging scenario space. Such nodes are root nodes in the space graph that are neither facts nor instantiated assumptions nor the justifying nodes that support the instantiated assumptions. That is, they are irrelevant to the actual case under investigation and, hence, should be removed. The removal procedure recursively examines the root nodes in the emerging scenario space and deletes such nodes. It terminates when each root node in the emerging scenario space corresponds to either a fact or an assumption, guaranteeing all the spurious nodes are removed.

IDENTITY RESOLUTION

To achieve a decision support system that is truly effective to battle real crimes within highly deceptive environment, a verification mechanism

is required to handle the identity problem, (Badia and Kantardzic 2005; Pantel 2006; Wang et al. 2005) thereby refining the quality of hypothesized scenarios. This analytical process is crucial as falsified identity has become the common denominator of all serious crime such as mafia trafficking, fraud, and money laundering. In particular to the September 11 terrorist attacks, tragic consequences could have been prevented to some extent if U.S. authorities had been able to discover the use of deceptive identity, e.g., multiple dates of birth and alias names. To such extent, many exiting verification systems (Branting 2003; Wang et al. 2006) that rely on the inexact search of textual attributes would fail drastically to recognize the unconventional truth like that between *Osama bin Laden* and *The Prince* (Hsiung et al. 2005).

Link Analysis Approach

The aforementioned dilemma may be overcome through link analysis that seeks to discover information based on the relations hidden in data about people, places, objects, and events. Link analysis techniques have proven effective for identity problems (Boongoen and Shen 2008; Hsiung et al. 2005; Pantel 2006) by exploiting link information observed from legitimate activities such as making use of mobile phones and financial systems. Essentially, to justify the similarity between entities (e.g., names) within a link network, several well-known algorithms, such as SimRank, (Jeh and Widom 2002) concentrate on the cardinality of joint neighbors to which they are linked (more details in Liben-Nowell and Kleinberg 2007).

Analogously, the Connected-Triple method (Klink et al. 2006) evaluates the similarity of objects in accordance with their overlapping social context. This measurement is based on the social network $G(V, E)$ in which objects of interest and their relations are represented by the set of vertices V and that of edges E , respectively. The similarity of any two objects can be estimated by counting the number of Connected-Triples of which they are a part. Formally, a Connected-Triple, $Triple = \{V_{Triple}, E_{Triple}\}$, is a subgraph of G containing three vertices $V_{Triple} = \{v_1, v_2, v_3\} \subset V$ and two edges $E_{Triple} = \{e_{v_1 v_2}, e_{v_1 v_3}\} \subset E$, with $e_{v_2 v_3} \notin E$. Figure 3 presents an example of a social network in which object A and object B are justified similar because there exists a Connected-Triple link connecting them together, $V_{Triple} = \{A, B, D\}$ and $E_{Triple} = \{e_{AD}, e_{BD}\}$, with $e_{AB} \notin E$.

Application to Scenario Refinement

Within a scenario generated by the method of the previous section, associations among information entities (e.g., persons, objects, and locations) can be perceived through n -nary ($n > 1$) predicates p (i.e.,

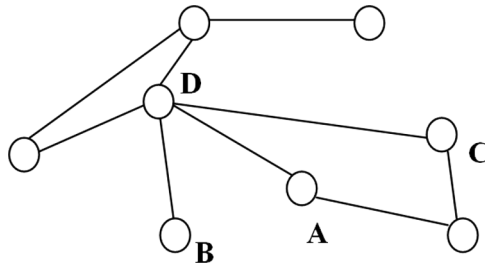


FIGURE 3 A social network with Connected-Triple.

a general notation for any p^s , p^a , and p^t in **If**, **Assuming**, and **Then** statements, respectively), instantiated through user inputs or the underlying abductive and deductive reasoning processes. To identify entity pairs that are likely to constitute alias references, a link network $G(V, E, W)$ (where V , E , and W represent a set of vertices, edges, and that of edge weights) is initially constructed from a given scenario S as follows:

- Let O and $p^k(o_1^k, \dots, o_m^k)$ be a set of entities included in the scenario S and a m -nary predicate specified over entities $o_1^k \dots o_m^k \in O$. In addition, an FCN \tilde{z}_{p^k} systematically expresses the quality of the predicate $p^k \in S$ to which it is attached. Each unique entity $o_i \in O$, which can be included in a single or multiple predicates, is represented as a vertex $v_i \in V$.
- A link between objects $o_i, o_j \in O$, which is denoted by an edge $e_{ij} \in E$ between the corresponding vertices $v_i, v_j \in V$, exists when they are part of q predicates ($q \geq 1$). In particular, each edge $e_{ij} \in E$ possesses $UT_{ij} = \{ut_{ij}^1, \dots, ut_{ij}^q\}$, a set of quality measures acquired from those q predicates, where $ut_{ij}^k = Re(\tilde{z}_{p^k}), \forall k = 1 \dots q$. Note that ut_{ij}^k is semantically represented by one of the fuzzy sets shown in Figure 1.
- Then, the weight $w_{ij} \in W$ of an edge $e_{ij} \in E$, where $w_{ij} \in \{0, VL, L, M, H, VH, 1\}$, can be estimated as

$$w_{ij} = \min_{\forall k \in UT_{ij}} ut_{ij}^k. \tag{9}$$

Having obtained the link network, the similarity $s(o_i, o_j)$ of any two entities $o_i, o_j \in O$ can be estimated from the number of their shared neighbors $|N_{o_i o_j}|$, where $N_{o_i o_j} \subset O$ is the set of objects o_x directly links to o_i and o_j , i.e., $e_{ix}, e_{jx} \in E, e_{ij} \notin E$. However, unlike the conventional Connected-Triple approach, the cardinality-based similarity measure $s(o_i, o_j)$ is adapted to weighted edges as

$$s(o_i, o_j) = \sum_{\forall n_x \in N_{o_i o_j}} SC(\min(w_{o_i n_x}, w_{o_j n_x})), \tag{10}$$

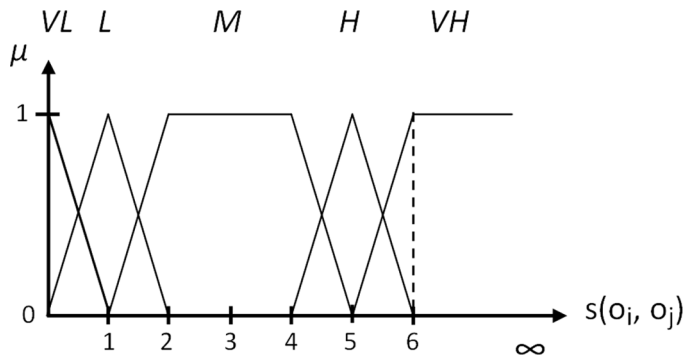


FIGURE 4 Definition of fuzzy sets for describing a similarity measure.

where $SC(f) \in [0, 1]$, $f \in \{0, VL, L, M, H, VH, 1\}$ denotes the score of each fuzzy set f , specifically exploited here to simplify the aggregation of collective measures: $SC(0) = 0$, $SC(VL) = \frac{1}{6}$, $SC(L) = \frac{2}{6}$, $SC(M) = \frac{3}{6}$, $SC(H) = \frac{4}{6}$, $SC(VH) = \frac{5}{6}$, $SC(1) = 1$. Other scoring schemes may be used, but this one is the simplest in implementation.

Fuzzy sets presented in Figure 4 are used to generate a coherently interpretable linguistic description of any similarity degree $s(o_i, o_j)$: $VL =$ Very Low, $L =$ Low, $M =$ Moderate, $H =$ High, and $VH =$ Very High. For instance, the similarity measure of 1.5 can be expressed as being $L(0.5)$ and $M(0.5)$, where membership degrees to corresponding fuzzy sets are given in brackets. Note that although the semantics of these descriptors are defined by domain experts in the current research, they can also be formulated from historical data using a machine learning technique. With this methodology, an investigator can efficiently differentiate and select the most likely case upon which the scenario is refined (see next section for an example).

ILLUSTRATIVE EXAMPLE

It is obvious that many explosive ingredients and liquids can be combined to create homemade liquid bombs. However, a lot of explosive chemicals can be concocted from some very common items, such as perfumes, drain cleaner, and batteries, and they are innocent in themselves. Also, the initial collected evidence may seem to be irrelevant with each other from the outset. This makes it very difficult for intelligence analysts to detect a plausible threat. This section shows how the proposed fuzzy compositional modeler and the link-based identity resolution can be used to generate and refine the plausible scenario space, given both vague and uncertain information in this application context.

Scenario Generation

Consider the following scenario: Assume that a suspect named Dave was trying to bring a bottle of coke on board on aircraft and a mobile phone is found on him. A few hours later the security control center found a small bag of hair dye in a suitcase under the name of Bob. Also, a light blue Nissan Micra was caught on the airport CCTV and Dave was alone in the car. With further investigation the police found out that this Nissan Micra belongs to Bob and the mobile phone is registered under the name of Bob as well. These pieces of evidence and some key SFs are listed in the Appendix. Note that because the evidence obtained from different sources may receive different certainty degrees, for example, the piece of evidence gathered by the CCTV camera record may receive a higher certainty degree than that described by a witness. To reflect this, the certainty degree (which may be initially assigned) of each piece of evidence is also provided.

Initial creation of the emerging scenario space is done by matching a given piece of evidence, say e_1 , against the knowledge base. It is obvious that MF_1 is first instantiated. This piece of evidence only involves boolean proposition, such that the fuzzy matching degree of e_1 and MF_1 is either 1 or 0. By applying Algorithm 1 to MF_1 , the results of Table 1 are obtained.

As indicated previously, to avoid generating unnecessary explanations, the modeler produces only the current best or the most plausible explanations in the first instance. With regard to the use of the modulus of a derived FCN for plausibility evaluation, only r_1 in MF_1 is instantiated for further development. As a result the first node, "Seen with(Bob,Hair dyes,Airport)", is added to the emerging scenario space with the FCN: $H + i1$. Given this and the instantiated MF_1 , backward chaining is performed from the existing node. This leads to the instantiated antecedences and assumption in MF_1 being added to the scenario space. According to Table 1, $\tilde{z}_{\text{antecedent},r_1} = 1 + i1$, it follows from Eq. (4) that

$$\begin{aligned} Re(\tilde{z}_{\text{antecedent},r_1}) &= \min(Re(\tilde{z}_{\text{Possess}(\text{Bob},\text{Hair_dyes}):True}), Re(\tilde{z}_{\text{Location}(\text{Airport}):True}), \\ &\quad Re(\tilde{z}_{\text{Substance}(\text{Hair_dyes}):True}), Re(\tilde{z}_{\text{Person}(\text{Bob}):True})), \\ Im(\tilde{z}_{\text{antecedent},r_1}) &= \min(Im(\tilde{z}_{\text{Possess}(\text{Bob},\text{Hair_dyes}):True}), Im(\tilde{z}_{\text{Location}(\text{Airport}):True}), \\ &\quad Im(\tilde{z}_{\text{Substance}(\text{Hair_dyes}):True}), Im(\tilde{z}_{\text{Person}(\text{Bob}):True})). \end{aligned}$$

TABLE 1 Results of Backward Chaining Due to e_1

Rule index	$Re(\tilde{z}_{\text{antecedent}})$	$Im(\tilde{z}_{\text{antecedent}})$
r_1	1	1
r_2	N/A	0

It is obvious that

$$\begin{aligned}
 Re(\tilde{z}_{\text{antecedent}_{r_1}}) = 1 &\iff \begin{cases} Re(\tilde{z}_{\text{Possess}(\text{Bob}, \text{Hair_dyes}): \text{True}}) = 1 \\ Re(\tilde{z}_{\text{Location}(\text{Airport}): \text{True}}) = 1 \\ Re(\tilde{z}_{\text{Substance}(\text{Hair_dyes}): \text{True}}) = 1 \\ Re(\tilde{z}_{\text{Person}(\text{Bob}): \text{True}}) = 1 \end{cases} \\
 Im(\tilde{z}_{\text{antecedent}_{r_1}}) = 1 &\iff \begin{cases} Im(\tilde{z}_{\text{Possess}(\text{Bob}, \text{Hair_dyes}): \text{True}}) = 1 \\ Im(\tilde{z}_{\text{Location}(\text{Airport}): \text{True}}) = 1 \\ Im(\tilde{z}_{\text{Substance}(\text{Hair_dyes}): \text{True}}) = 1 \\ Im(\tilde{z}_{\text{Person}(\text{Bob}): \text{True}}) = 1 \end{cases}
 \end{aligned}$$

Therefore the FCNs associated with the newly added scenario nodes can be derived.

Similarly, the remaining pieces of evidence are used to match with the knowledge base sequentially. Note that, for those existing nodes in the emerging scenario space, their associated FCNs are updated by using Eq. (8). As a result, after the backward-chaining and forward-chaining phase, the generated scenario space is depicted in Figure 5.

Identity Resolution

According to the scenario given in Figure 5, the link network (shown in Figure 6) is achieved by following the method set out in the previous section. In particular, several entity pairs are identified with non-zero similarity values, thus forming a candidate list of duplicates. For instance,

- $s(\text{Bob}, \text{Dave}) = \text{Low}(0.5), \text{Moderate}(0.5)$
- $s(\text{Mobile phone}, \text{Nissan}) = \text{Low}(0.67), \text{Moderate}(0.33)$
- $s(\text{Hair dyes}, \text{Nissan}) = \text{Very Low}(0.33), \text{Low}(0.67)$
- $s(\text{Coke}, \text{Mobile phone}) = \text{Very Low}(0.5), \text{Low}(0.5)$
- $s(\text{Coke}, \text{Nissan}) = \text{Very Low}(0.5), \text{Low}(0.5)$

With fuzzy linguistic similarity descriptors, an investigator can conveniently perceive and differentiate among candidate pairs. The higher the similarity value, the greater the possibility of the entity pair being duplicates becomes. Intuitively, only the pair of (Bob, Dave) is applicable because others include incompatible entity types (e.g., Coke and Nissan), and they are justified with lower similarity degrees. Hence, Bob and Dave are accepted to be different references to the same person. This new finding is then used to refine the existing scenario space of Figure 5 by

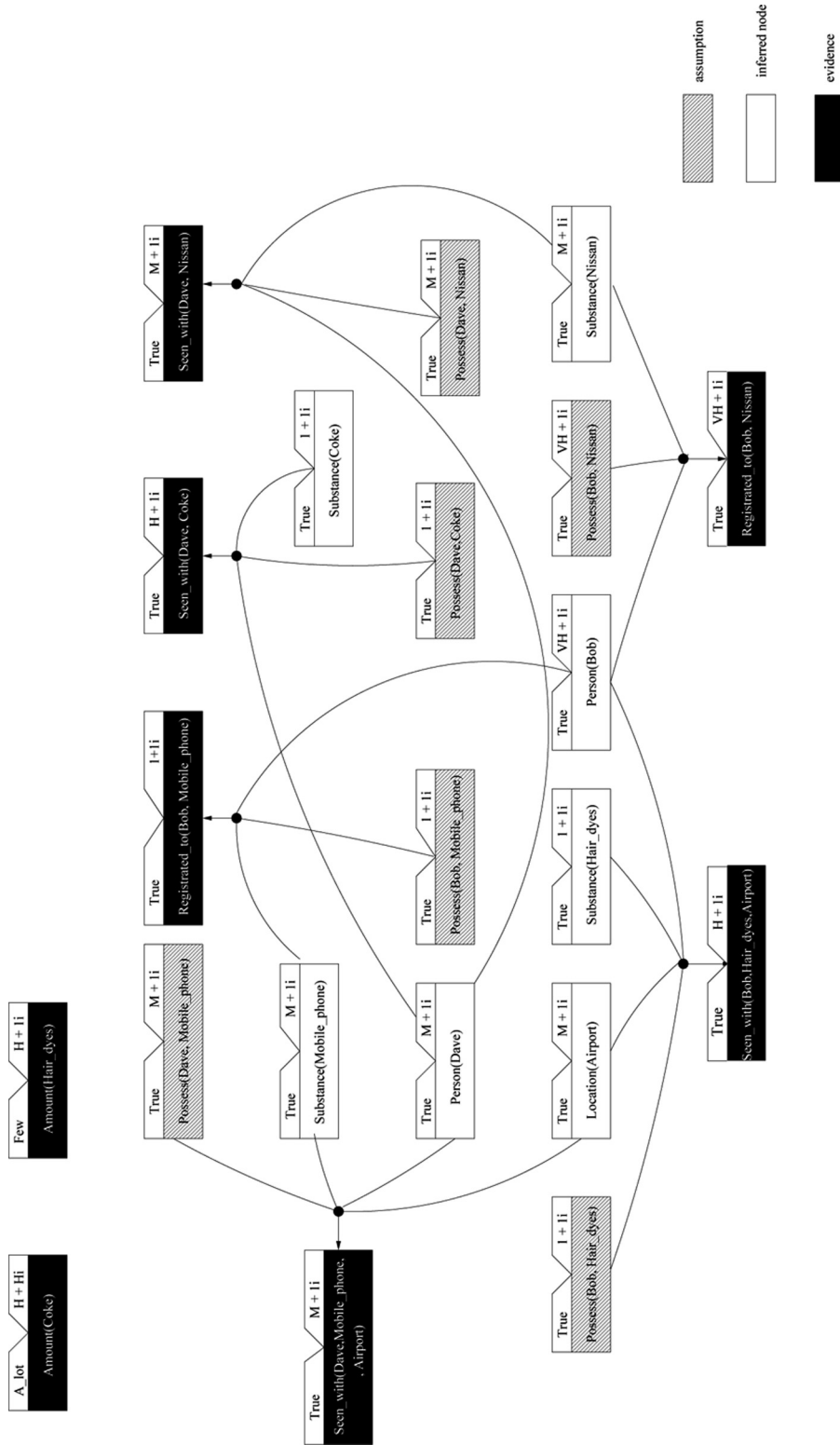


FIGURE 5 Generated scenario space.

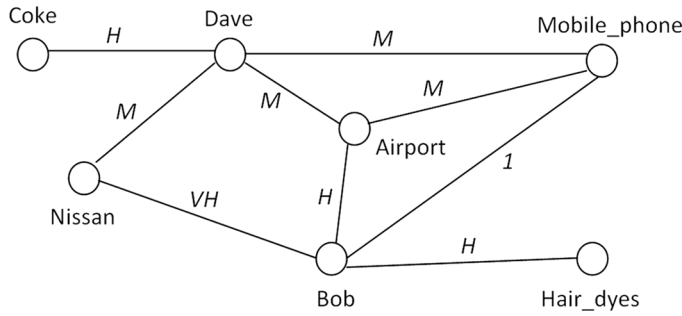


FIGURE 6 A link network with fuzzy linguistic weights.

activating MF_4 . The forward chaining procedure is then performed and the results are listed in Table 2.

According to the modulus of the FCN obtained for each matched proposition instance, the one associated with r_1 in MF_4 outperforms those associated with the rest. The revised scenario space is depicted in Figure 7. It is noteworthy that a new and interesting hypothesis of *preparing a liquid bomb* emerges as the result of this identity fusion. In addition to the capability to discover an unforeseen scenario, the link-based approach is able to help the CM modeler scaling up to a large problem, in which instantiated instances can be efficiently managed.

CONCLUSION

This article presents a novel approach to generate and refine plausible crime scenarios from available evidence. It has two significant advantages: i) the ability to represent and reason with information at varying degrees of precision and ii) the ability to enhance the quality of generated scenarios through the resolution of duplicated references. Initially, a knowledge representation formalism is introduced to represent both vague and uncertain information. In particular, FCN propagation algorithms are developed and integrated to the conventional abductive modeler (Shen et al. 2006) to generate plausible scenarios. Having obtained scenarios,

TABLE 2 Results of Forward-Chaining Propagation Within MF_4

Rule index	$Re(\tilde{z}_c)$	$Im(\tilde{z}_c)$
r_1	H	H
r_2	VL	VL
r_3	VL	VL
r_4	H	L

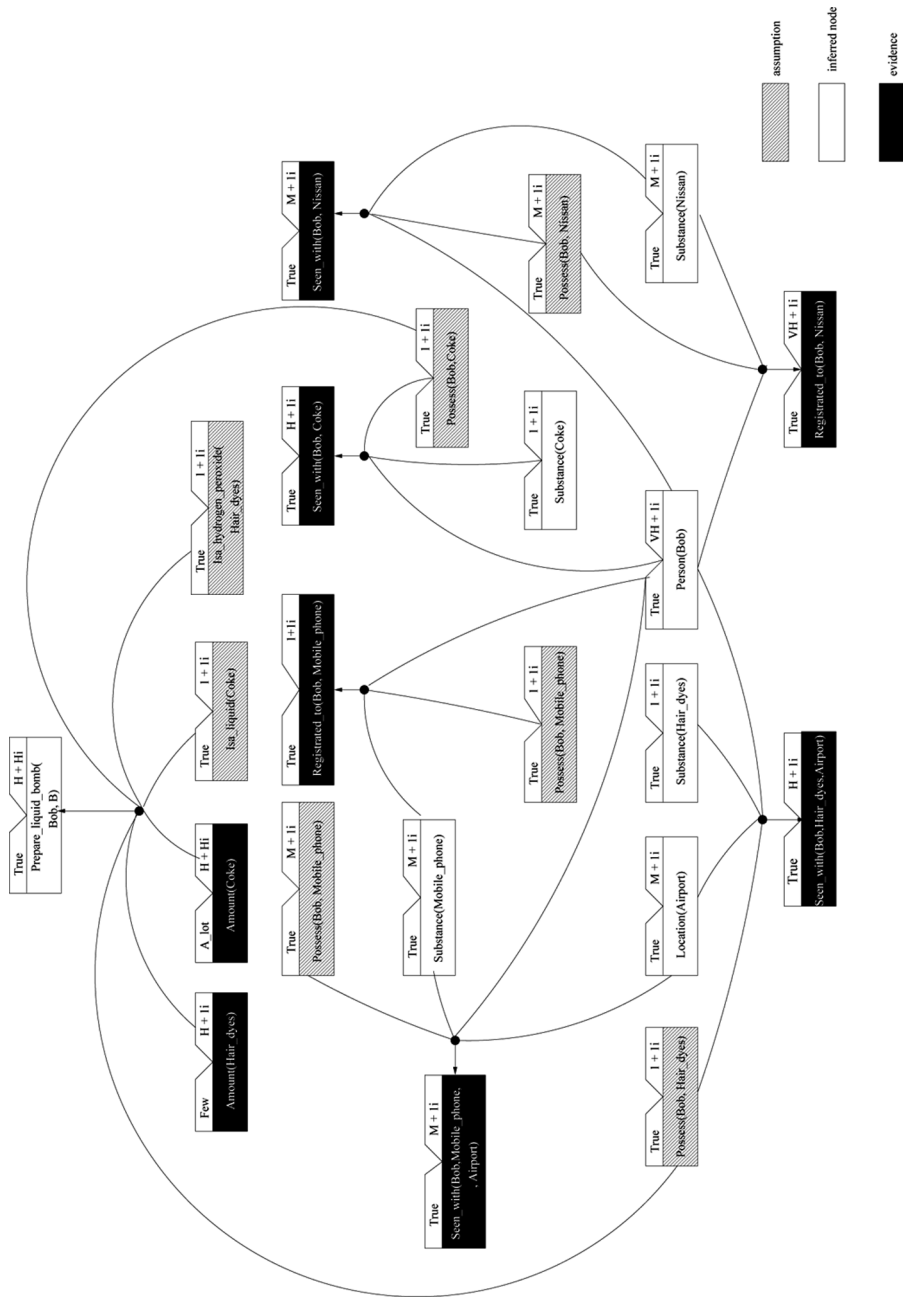


FIGURE 7 Updated scenario space.

a link-based method is used to identify potential duplicates (i.e., aliases), which are consequently exploited to refine the underlying scenarios.

To extend its applicability, an ontological methodology may be introduced such that the knowledge-based decision support system becomes compatible with a variety of information entities and their relations. Essentially, the similarity among entities, estimated from the ontology, can improve the effectiveness of both scenario generation (semantic matching) and identity resolution (filtering out candidates with dissimilar-typed entities). Also, a more general constraint satisfaction mechanism suitable for constraining variables modified by FCNs would help to improve the generality of the proposed approach. In particular, this may avoid the need to prefix just one common quantity space to which the real and imaginary parts of any FCN may belong.

APPENDIX KEY SFs IN SAMPLE KNOWLEDGE BASE

Define **fuzzyvariable**{

Name = Amount

Universe of discourse: [0, 1]

Unit = none

Cardinality of partition = 5

Quantity space:

$$fs_1 = [0, 0, \frac{1}{n-1}]$$

...

$$fs_i = [\frac{i-2}{n-1}, \frac{i-1}{n-1}, \frac{i}{n-1}]$$

...

$$fs_n = [\frac{n-2}{n-1}, 1, 1]$$

Names of fuzzy sets = {none, few, several, many, a lot}

Unifiability = Amount(X)}

If {Person(P), Substance(X), Location(L)}

Assuming {Possess(P, X)}

Then {Seen with(P, X, L)}

Distribution Seen with(P, X, L) {

r1: true, true, true, true → true:H,

r2: true, true, true, true → false:VL} (MF_1)

If {Person(P), Substance(X)}

Assuming {Possess(P, X)}

Then {Seen with(P, X)}

Distribution Seen with(P, X) {

r1: true, true, true \rightarrow true:H,
 r2: true, true, true \rightarrow false:VL} (MF_2)

If {Person(P), Substance(X)}
Assuming {Possess(P, X)}
Then {Registered to(P, X)}
Distribution Registered to(P, X) {
 r1: true, true, true \rightarrow true:VH,
 r2: true, true, true \rightarrow false:VL} (MF_3)

If {Amount(X), Amount(Y), Possess(P, X), Possess(P, Y)}
Assuming {Isa liquid(X), Isa hydrogen peroxide(Y)}
Then {Prepare liquid bomb(P, B)}
Distribution Prepare liquid bomb(P, B) {
 r1: many, few, true, true, true, true \rightarrow true:VH,
 r2: many, few, true, true, true, true \rightarrow false:VL,
 r3: few, many, true, true, true, true \rightarrow true:VL,
 r4: few, many, true, true, true, true \rightarrow false:VH} (MF_4)

Collected Evidence

- e_1 : Seen with(Bob, hair dyes, airport) = true with certainty degree:H
- e_2 : Registered to(Bob, Nissan) = true with certainty degree:VH
- e_3 : Seen with(Dave, mobile phone, airport) = true with certainty degree:M
- e_4 : Registered to(Bob, mobile phone) = true with certainty degree:l
- e_5 : Seen with(Dave, coke) = true with certainty degree:H
- e_6 : Seen with(Dave, Nissan) = true with certainty degree:M
- e_7 : Amount(coke) = a lot with certainty degree:H
- e_8 : Amount(hair dyes) = few with certainty degree:H

REFERENCES

- Badia, A., and M. M. Kantardzic. 2005. Link analysis tools for intelligence and counterterrorism. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics*, 49–59. Atlanta, Georgia: Springer Berlin.
- Boongoen, T., and Q. Shen. 2008. Detecting false identity through behavioural patterns. In *Proceedings of International Crime Science Conference*, London, UK: Willan Publishing.
- Boongoen, T., Q. Shen, and C. Price. Disclosing false identity through hybrid link analysis. *AI and Law*, In press (DOI: 10.1007/510506-010-9085-9)
- Branting, K. 2003. A comparative evaluation of name matching algorithms. In *Proceedings of International Conference on AI and Law*, 224–232. Edinburgh, Scotland: ACM.
- Cooman, G. D. 2005. A behavioural model for vague probability assessments. *Fuzzy Sets Syst.* 154:305–358.

- Fu, X., and Q. Shen. 2009. A novel framework of fuzzy complex numbers and its application to compositional modelling. In *Proceedings of the 2009 IEEE International Conference on Fuzzy Systems*, 536–541. ICC, Jeju Island, Korea: IEEE.
- Fu, X., Q. Shen, and R. Zhao. 2007. Towards fuzzy compositional modelling. In *Proceedings of the 16th International Conference on Fuzzy Systems*, 1233–1238. London, UK: IEEE.
- Halliwell, J., and Q. Shen. 2009. Linguistic probabilities: theory and application. *Soft Computing* 13:169–183.
- Halliwell, J., J. Keppens, and Q. Shen. 2003. Linguistic bayesian networks for reasoning with subjective probabilities in forensic statistics. In *Proceedings of International Conference on Artificial Intelligence and Law*, 42–50. Edinburgh, Scotland: ACM.
- Hsiung, P., A. Moore, D. Neill, and J. Schneider. 2005. Alias detection in link data sets. In *Proceedings of International Conference on Intelligence Analysis*, McLean, VA: International Conference on Intelligence Analysis.
- Jeh, G., and J. Widom. 2002. Simrank: A measure of structural-context similarity. In *Proceedings of International Conference on Knowledge Discovery and Data Mining*, 538–543. Edmonton, Alberta: ACM.
- Keppens, J., and Q. Shen. 2001. On compositional modelling. *Knowledge Engineering Review* 16:157–200.
- Keppens, J., Q. Shen, and B. Schafer. 2005. Probabilistic abductive computation of evidence collection strategies in crime investigation. In *Proceedings of International Conference on Artificial Intelligence and Law*, 215–224. Bologna, Italy: ACM.
- Klink, S., P. Reuther, A. Weber, B. Walter, and M. Ley. 2006. Analysing social networks within bibliographical data. In *Proceedings of the 17th International Conference on Database and Expert Systems Applications*, 234–243. Krakow, Poland: Springer Verlag.
- Liben-Nowell, D., and J. Kleinberg. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* 58:1019–1031.
- Pantel, P. 2006. Alias detection in malicious environments. In *Proceedings of AAAI Fall Symposium on Capturing and Using Patterns for Evidence Detection*, Washington, DC, 14–20.
- Shen, Q., J. Keppens, C. Aitken, B. Schafer, and M. Lee. 2006. A scenario-driven decision support system for serious crime investigation. *Law, Probability and Risk* 5:87–117.
- Wang, G. A., H. Atabakhsh, T. Petersen, and H. Chen. 2005. Discovering identity problems: A case study. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics*, Atlanta, 368–373.
- Wang, G. A., H. Chen, J. J. Xu, and H. Atabakhsh. 2006. Automatically detecting criminal identity deception: an adaptive detection algorithm. *IEEE Transactions on Systems, Man and Cybernetics, Part A* 36:988–999.