Aberystwyth University

Fuzzy compositional modeling

Fu, Xin; Shen, Qiang

Fuzzy Compositional Modeling

Xin Fu, *Student Member, IEEE*, and Qiang Shen

Abstract—Automated modeling refers to automatic (re-) formulation of alternative system models that embody the simplification, abstraction, and approximation of knowledge and data for a given task. This technique is highly desirable for effective problem solving in many application domains. Over the past two decades, compositional modeling (CM) has established itself as a leading approach in automated modeling. CM is a framework to construct system models by composing generic and reusable model fragments (MFs) selected from a knowledge base. However, the existing work mainly concerns the knowledge and data that are represented by crisp and precise information. Little work has been carried out to explore its potential to deal with uncertain environments. This paper presents an innovative framework of fuzzy compositional modeling (FCM) to develop such work. The proposed approach is capable of representing and reasoning with a wide range of inexact information. An innovative notion of fuzzy complex numbers (FCNs) is developed in an effort to enable synthesis of consistent scenario descriptions from imprecise MFs. This paper also introduces the modulus of FCNs to constrain the resulting scenario descriptions. The usefulness of this study is illustrated by means of an example to construct possible scenario descriptions from given evidence, which is in support of crime investigation.

Index Terms—Compositional modeling (CM), crime investigation, fuzzy complex numbers (FCNs).

I. INTRODUCTION

EFFECTIVE problem solving often requires the automatic construction of computational models that are both adequate and efficient for a given task. However, automated modeling is a complex problem: Generation of appropriate models requires the consideration of not only the aims and intentions of the task specification but user preferences as well. Nevertheless, several automated modelers have been proposed in the literature, which specialize in performing different tasks and supported with various fundamental architectures. Among these is the work of compositional modeling (CM) [7], [16], which is the knowledge-based approach, and has established itself as a leading tool to synthesize and store plausible-scenario space effectively and efficiently in many problem domains (e.g., those of physics [7], ecology [17], criminology [27], and social science [19]).

While existing CM work has demonstrated the potential of the underlying approach, its use is restricted in real-world applications, as it assumes that the model fragments (MFs) within the knowledge base (KB) are expressed by precise and crisp information. In coping with uncertainty, a set of numerically specified probability distributions is typically employed by each MF to represent the likelihood of its associated outcomes. However, many problem domains involve imprecise, incomplete, and empirical knowledge and data. The degree of precision of available information may vary greatly, subject to different perceptions and judgement of users. In particular, assessment of likelihood reflects the expertise of the user and is often expressed in qualitative terms, i.e., verbally or diagrammatically. In fact, the use of seemingly very precise numeric probabilities typically suffers from an inadequate degree of accuracy [14].

Dealing with inexact knowledge and data captured in a variety of forms has thus become an increasingly important issue in CM. Conceptually, inexact information may be classified into the following four general categories, which are of particular interest to this study.

1) *Vagueness:* It arises due to a lack of sharp distinctions or boundaries between pieces of information (e.g., Bob is *tall*, not *medium*). Often, a vague proposition is modeled by a fuzzy set that identifies a soft constraint on a set of elements. Instead of using a crisp partition, an element of a fuzzy set is allowed to satisfy the soft constraint to a certain degree.

2) *Uncertainty:* It depicts the reliability or confidential weight of a given piece of information stated in a proposition. Due to the involvement of uncertainty, it is difficult to state the exact truth of a given statement. In the literature, two types of uncertainty are often referred to as randomness and fuzziness [20], [29]. Probability theory is typically employed to model randomness, while the possibility theory [33] is well established to deal with fuzziness (where a fuzzy-set membership function is interpreted as a possibility distribution). This work mainly concerns the latter type of uncertainty, which is represented by a numerical value (e.g., The suspect overpowers the victim, with a certainty degree of 0.7).

3) *Both vagueness and uncertainty:* This means that information of type 1 and that of type 2 coexist (e.g., The amount of collected fiber is *a_lot*, with a certainty degree of 0.7).

4) *Both vagueness and uncertainty with the latter also expressed in vague terms:* Instead of using numerical values, the uncertainty is described by a linguistic term (e.g., The amount of collected fiber is *a_lot*, with a certainty degree of *very_likely*).

Much work has been developed to support reasoning with inexact knowledge and data [18], [21]. Although the application problems and the problem-solving approaches taken may be rather different, the existing techniques all aim to integrate the underlying distinct pieces of inexact information into a global
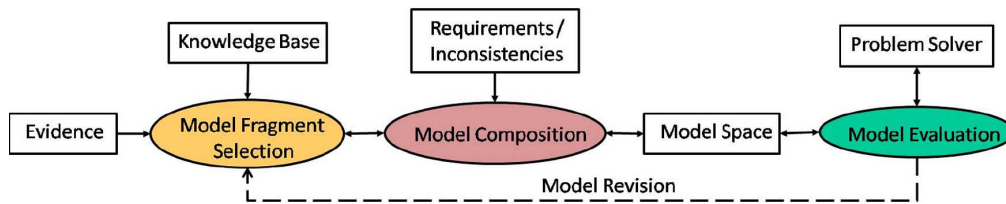
Fig. 1.   Architecture of CM.

measure. However, in performing such an integration, the underlying semantics associated with different information components may be destroyed. It is of great interest and potentially beneficial to establish a new mechanism that will maintain the associated semantics when reasoning with inexact knowledge and data. This paper proposes a novel fuzzy-CM framework in an attempt to instantiate and compose MFs into consistent scenario descriptions given the aforementioned types of information. In order to achieve this, an innovative notion of fuzzy complex numbers (FCNs) [10] is employed to provide an effective and efficient representation and inference of 2-D inexactness (i.e., vagueness and uncertainty) conjunctively and explicitly.

Note that the term FCN is not new; the concept of complex numbers has been proposed in the literature. In particular, a form of *FCNs* has been defined in [2] as a mapping from the complex plane to the interval $[0, 1]$. Such an FCN is, therefore, simply a conventional type-1 fuzzy set. Work on the differentiation and integration of this type of *FCN* has been proposed in [3] and [4] as well, with more advanced follow-on research reported in [23], [24], [31], [32], and [35]. Recently, in combining fuzzy complex analysis and statistical learning theory, important theorems (of a learning process) based on fuzzy-complex random samples were developed [13]. This work further establishes interesting properties of the so-called *rectangular FCNs*, which are special types of *FCNs*, as proposed in [2]. Another interesting development is the notion that relates real complex numbers to fuzzy sets [25]. It introduces a new type of set, which is called *complex fuzzy sets*, to allow the membership value of a standard fuzzy set to be represented using a classical complex number. However, as discussed in [25], it may be difficult to acquire an understanding for the use of complex-valued memberships. Despite this obstacle, work has continued along this theme of research. This is evident in that *complex fuzzy sets* have been integrated with propositional logic to construct fuzzy-reasoning systems [6].

Existing work related to the concept of FCNs is nevertheless framed by either giving conventional complex numbers a real-valued membership or assigning a fuzzy-set element to a complex number as its membership value. These are rather different from what is proposed in this paper, where both the real and imaginary values of an FCN are, in general, themselves fuzzy numbers, each with an embedded semantic meaning. In particular, the calculus of such FCNs over arithmetic and propositional relations is purposefully developed to support scenario-model synthesis from MFs, and the modulus of FCNs is introduced to constrain the emerging scenario descriptions. Of course, the underlying development of this new approach to FCNs is general and may be further adapted to other application problems.

From the CM point of view, this work not only provides a more flexible knowledge representation formalism, but also enhances the capability of CM in handling a variety of inexact information. In order to achieve this, the following technical inventions with respect to CM are developed. First, due to the involvement of vague information, a precise and certain match between the available evidence and the MFs, in general, cannot be expected. The boolean retrieval approach used in existing CM work fails to return any MFs that partially match the available information. Hence, a fuzzy mechanism is proposed to retrieve those MFs that involve no exact match and are most likely to be relevant to the collected evidence. Second, a heuristic method for model composition, in conjunction with its associated inference mechanism is developed, for the first time, to dynamically synthesize the retrieved (fuzzy) MFs in order to construct consistent plausible scenarios. In particular, two algorithms are designed to deal with the backward and forward propagation of the involved inexact knowledge and data. Further, several filtering methods are also introduced to refine the emerging scenario descriptions.

The rest of the paper is organized as follows. Section II introduces the theoretical background that provides a basis for this work. A structured knowledge-representation formalism is described in Section III to represent and store various types of inexact information within CM. This is based on the initial work of Fu *et al.* [11]. Section IV presents the novel fuzzy compositional modeler that incorporates an inference mechanism to match, compose, and propagate inexact information within the model composition process. This is achieved by substantially extending and combining the seminal ideas proposed in [9] and [10]. An illustrative example is then given in Section V, showing the utility and usefulness of this approach in providing decision support in crime investigation. Finally, Section VI concludes the paper with future work pointed out.

II. OUTLINE OF COMPOSITIONAL MODELING

A. Compositional Modeling

A compositional modeler constructs a mathematical or conceptual model based on certain initial knowledge and data through computational means [16]. The generic architecture of CM is depicted in Fig. 1. Given a piece of evidence (which may come in a variety of forms, e.g., an observation, a query, or a description of the expected behavior of the resulting model) and a predefined KB, the task of CM is designed to create models that represent the most useful and coherent plausible scenarios that may explain the obtained evidence.

The process of CM starts with *MF selection*, i.e., by matching the evidence to the relevant referents in the KB, thereby

identifying a subset of those available MFs. This module has an embedded inference mechanism that instantiates, possibly partially, the relevant MFs with known information. After this, the *model composition* subprocess assembles the selected MFs into plausible-scenario models by taking into consideration various requirements and specifications of (in-)consistencies. Conceptually, this subprocess can be viewed as a search step, where the goal is to select an appropriate and consistent scenario model from a space of plausible models created by the retrieved MFs and the initial evidence. In implementing this module, a number of techniques may be employed. In particular, a *consistency-checking mechanism* can be adopted to ensure that the emerging model is compatible with the entire environment in which the models are to be built. In addition, *causal-ordering techniques* may be used to establish cause-and-effect relations over the variables in the models.

Often, there are more than one plausible-scenario model that may be generated, although they are not always equally suitable for the problem at hand. The quality of a model depends on many aspects, including the necessity of the components, the quality of the given KB, and the adequacy of the included assumptions. All candidate models are therefore evaluated in the *model evaluation* module with respect to user-specified criteria, such as simplicity, completeness, or other preferences. The resulting most-appropriate model is passed on to the *problem-solver* module to implement the actual model-based application system. During the *model evaluation* and problem-solving phases, new pieces of information that conflict the current inference and assumptions may be derived. For example, a certain assignment of one variable may conflict with the observation of its environment or certain variables may be out of the scope of the operating ranges assumed by the model. Such contradictions are then fed back to the *MF-selection* module to revise the models. Note that this generic architecture of CM will be further extended in Section IV to support fuzzy-CM task.

B. Knowledge Representation

The KB of the CM consists of a number of reusable MFs that represent generic relationships between domain concepts and their states for a certain partial scenario. Different MFs can describe different partial scenarios or can be different descriptions of a common partial scenario. The generality of MFs is inherited from the first-order representation in which only variables appear in an MF, such that the variables can be instantiated by being assigned certain values.

Definition 1: An MF is a tuple $\langle \mathbf{X}, \mathbf{Y}, \mathbf{H}, \mathbf{R} \rangle$, which is described as follows.

> IF $\mathbf{X}$
> *Assuming* $\mathbf{H}$
> THEN $\mathbf{R}(\mathbf{X}, \mathbf{H}, \mathbf{Y})$

1) $\mathbf{X} = \{x_1, \ldots, x_n\}$ is a set of antecedent predicates, which refers to already-identified objects of interest in the partial scenario.
2) $\mathbf{H} = \{h_1, \ldots, h_q\}$ is a (possibly empty) set of assumptions, which refers to those pieces of information that are unknown or currently cannot be inferred from others, but they may be presumed to hold for the sake of performing hypothetical reasoning.
3) $\mathbf{Y} = \{y_1, \ldots, y_m\}$ is a set of consequent predicates, which describe the consequences when the conditions and assumptions hold, including pieces of new knowledge or relations that are derived from the hypothetical reasoning.
4) $\mathbf{R} = \{r_1, \ldots, r_f\}$ is a set of relations imposed by this MF. It describes constrains over the objects/predicates in $\mathbf{X}, \mathbf{Y}$, and $\mathbf{H}$.
5) The IF statement describes the required conditions for a partial scenario to become applicable. These conditions must be at least true to a certain degree or logical consequences of other instantiated MFs.
6) The *Assuming* statement indicates the reasoning environment, which specify the uncertain events and states that are presumed in a partial scenario description.
7) The THEN statement concludes the consequent when the conditions and presumed assumptions hold.

C. Significance of Compositional Modeling

The most important property of CM is its ability to automatically construct many variations of a given problem from a relatively small KB, which is possible because the constituent parts of different scenarios are not normally unique to any one specific scenario, i.e., there are potentially many scenarios that possess common or similar properties locally and globally. These partial scenarios, including scenario elements and their relationships, can therefore be modeled as generic and reusable MFs and only need to be recorded once in the KB.

Second, a compositional modeler can compose a range of candidate models for a given task specification. As such, CM provides an efficient and compact means to represent and store modeling knowledge of different types of perspective. With the help of CM, users may not only obtain the models that they expect but also discover those plausible models that have never been considered.

Another significant property of CM is that it allows hypothetical reasoning to be performed by assembling MFs to form a variety of scenario models under different sets of model assumptions. This is often required as the available data and knowledge may be insufficient to build a complete model description (particularly at the initial stage of model building). CM possesses this property due to the default utilization of modeling assumptions.

III. KNOWLEDGE REPRESENTATION UNDER INEXACTNESS

Effective knowledge representation is essential to the development of CM framework. This section focuses on the creation of a structured knowledge-representation scheme that is capable of storing and managing inexact knowledge and data. It involves two conceptually distinct aspects: 1) introduction of a KB that consists of a weighted taxonomy and generic fuzzy MFs and 2) the definition of an innovative notion of FCNs that facilitates the propagation of vagueness and uncertainty within the process of model composition.
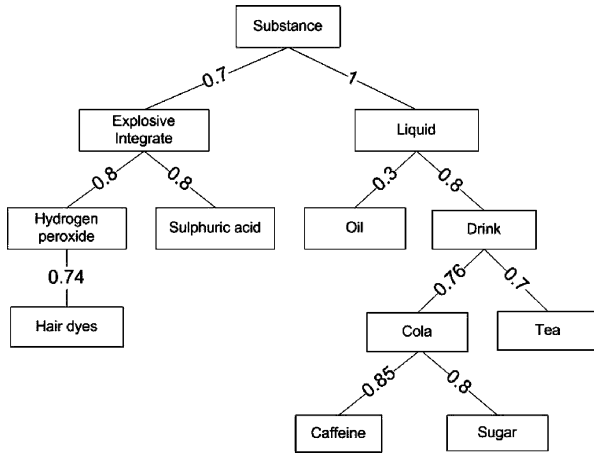
Fig. 2.   Portion of a weighted taxonomy.

A. Weighted Taxonomy

In this work, a weighted taxonomy is employed to represent the collection of concepts and their relationships within a given problem domain. Many concepts may exist that share structural similarities, those concepts that share aspect(s) in common or are highly relevant are grouped into the same class. For example, Fig. 2 shows part of such a weighted taxonomy for the problem domain of crime investigation.

In a given taxonomy, not only nodes but also arcs can carry semantic information. A weight attached to each arc indicates the degree of relevance between a child node and its parent. In other words, it expresses to what extent the child can be classified as an element of the class representing its parent node. The higher weight a concept receives, the more common features it inherits from its parent. These numeric weights are assigned by domain experts. Note that, in general, such weights may be given in linguistic terms. However, as an initial work on fuzzy CM, they are assumed to be expressed numerically only to keep this development focussed on the essentials. Note also that since a concept may belong to different categories, the sum of the arc weights of subtype concepts does not have to be equal to 1. Each weight can be assigned independently and subjectively, without taking into account its sibling nodes. Furthermore, the weighted taxonomy is independent of whether the concepts themselves are fuzzy.

B. Fuzzy-Model Fragments

In this work, fuzzy-set theory is employed to represent vague/ill-defined concepts and quantities in MFs. Such concepts and quantities are defined as fuzzy variables and their values are fuzzy sets. The representation of fuzzy MFs involves the following two aspects: 1) identification and definition of fuzzy variables in a generic sense and 2) specification of fuzzy constraints over involved fuzzy/boolean variables.

1) Fuzzy Parameters: Fuzzy variables are indicated by means of the keyword *fuzzyvariable*.

Definition 2: A fuzzy variable is defined by specifying the following parameter fields.

1) *Name:* It is a constant that uniquely identifies the fuzzy variable.
2) *Universe of discourse:* It is the domain of the fuzzy variable.
3) *Unit:* It represents the variable's physical dimension.
4) *Cardinality of partition:* It represents the number of fuzzy sets that jointly partition the universe of discourse. It is encoded by a positive integer.
5) *Quantity space [28]:* It is the collection of all the membership functions, which define the fuzzy sets that jointly cover the partitioned domain.
6) *Name of fuzzy sets:* It represents the symbolic label of each fuzzy set in the quantity space.
7) *Unifiability:* It is the declaration of a unifiable property of the variable, which is specified by a predicate.

The following example shows a fuzzy variable, which is named *Chance*, and its associated quantity space, as illustrated in Fig. 6. Let us define *fuzzyvariable*{

Name: Chance   *Universe of discourse*: $[0, 1]$
Unit: none   *Cardinality of partition*: $n = 7$
Quantity space:

$$fs_1 = \left[0, 0, \frac{1}{n-1}\right]$$
$$\cdots$$
$$fs_i = \left[\frac{i-2}{n-1}, \frac{i-1}{n-1}, \frac{i}{n-1}\right]$$
$$\cdots$$
$$fs_n = \left[\frac{n-2}{n-1}, 1, 1\right]$$

Name of fuzzy sets: {0, VL, L, M, H, VH, 1}
Unifiability: Chance(X)}

2) Fuzzy Constraints: In CM, knowledge is normally expressed as constraints or relations that must be obeyed by certain variables. For example, velocity and duration relations often appear in physical systems; length and angle relations often appear in spatial reasoning systems. In this work, the relations between domain elements are represented in a form similar to the style of production rules but involve much more general contents. However, variables under different situations may behave differently. Thus, within such generic rules, a set of rule instances may also be included, which are interchangeably termed as candidate assignments. These rule instances represent how certain the corresponding (possibly partially) instantiated relationships hold.

Definition 3: A fuzzy MF is adapted to possess the following form:

IF   **X**   *Assuming* **H**   THEN $\mathbf{X}, \mathbf{H} \rightarrow \mathbf{Y}$
Distribution **Y**
$\{A_{j_1}^1, \ldots, A_{j_n}^n, C_{i_1}^1, \ldots, C_{i_q}^q \rightarrow B_{j_{n+1}}^1, \ldots, B_{j_{n+m}}^m : p\}$

where $\mathbf{X} = \{x_1, \ldots, x_n\}$, $\mathbf{Y} = \{y_1, \ldots, y_m\}$, and $\mathbf{H} = \{h_1, \ldots, h_q\}$ are the antecedent, consequent, and assumption variables, respectively. Within the *Distribution* specification, the left-hand side of the "implication" sign in each propositional

instance is a combination of value pairs involving the values of antecedent and assumption variables, and the right-hand side indicates the corresponding possible outcome if the MF is instantiated, where

$$A_{j_l}^l \in D_{x_l} = \{A_1^l, A_2^l, \ldots, A_{k_l}^l\}$$
$$j_l = 1, 2, \ldots, k_l, \qquad l = 1, 2, \ldots, n$$
$$\cdots$$
$$C_{j_{n+l}}^l \in D_{z_l} = \{C_1^l, C_2^l, \ldots, C_{k_{n+l}}^l\}$$
$$j_{n+l} = 1, 2, \ldots, k_{n+l}, \qquad l = 1, 2, \ldots, q$$
$$\cdots$$
$$B_{j_{n+q+l}}^l \in D_{y_l} = \{B_1^l, B_2^l, \ldots, B_{k_{n+q+l}}^l\}$$
$$j_{n+q+l} = 1, 2, \ldots, k_{n+q+l}, \qquad l = 1, 2, \ldots, m.$$

The sets D_{x_l}, D_{z_l}, and D_{y_l} are the domains of corresponding variables and $A_{j_l}^l$, $C_{j_{n+l}}^l$, and $B_{j_{n+q+l}}^l$ are the domain elements. In each rule instance, p indicates the certainty degree of its corresponding outcome if the MF is instantiated.

The existing work on CM either does not handle uncertainty or requires numerical values to quantify the probability of a consequence's occurrence. However, such subjective assessment is usually the product of barely articulate intuition. This seemingly numerically precise expressions may cause loss of efficiency, accuracy, and transparency [5], [14]. Often, experts may be unwilling or simply unable to suggest a numerical probability. Therefore, the work developed here intends to capture the vagueness of the probability distribution in terms of subjective certainty degrees. Rather than using numerical values, fuzzy numbers or their corresponding linguistic terms are introduced to represent such subjective certainty degrees.

For simplicity, when describing a rule instance, the assumptions are treated as part of the antecedent due to their logical equivalence, i.e., a rule instance in an MF is of the form

$$\text{IF } x_1 \text{ is } A_{j_1}^1 \oplus \cdots \oplus x_n \text{ is } A_{j_n}^n$$
$$\text{THEN } y_1 \text{ is } B_{j_{n+1}}^1 \otimes \cdots \otimes y_m \text{ is } B_{j_{n+m}}^m \qquad (1)$$

where the operator $\otimes$ in the consequence is restricted to logical conjunction in this paper, but the $\oplus$ in the antecedence denotes either a conjunctive or disjunctive operator. Further, this rule instance can be decomposed to multiple simpler rules, each involving a single consequence, and can equivalently be written as

$$\text{IF } x_1 \text{ is } A_{j_1}^1 \oplus \cdots \oplus x_n \text{ is } A_{j_n}^n, \text{ THEN } y_1 \text{ is } B_{j_{n+1}}^1$$
$$\cdots$$
$$\text{IF } x_1 \text{ is } A_{j_1}^1 \oplus \cdots \oplus x_n \text{ is } A_{j_n}^n, \text{ THEN } y_m \text{ is } B_{j_{n+m}}^m.$$

Hence, in the remainder of this work, only those rules with single consequence will be considered. The following MF illustrates the concepts and applicability of fuzzy parameters and constraints:

IF {height(S), height(V)}
Assuming {attempted_to_kill(S,V)}
THEN {difficulty(overpower(S,V))}

Distribution difficulty(overpower(S,V)){
r_1: tall, short, true → easy: VH
r_2: short, tall, true → easy: VL}.

This MF describes a general relation between the height of two people involved in a fight and the difficulty for one to overpower the other. Here, the *height* is modeled as a fuzzy variable that takes values from the quantity space of $Q = \{very\_short,$ *short, average, tall, very_tall*}, and *difficulty* is another fuzzy variable with possible values of *easy, average*, and *difficult*. Specifically, the MF covers two rule instances, which indicate that if suspect S is *tall*, while victim V is *short*, and S indeed attempted to kill V, then S stands a *Very_High* (VH) chance to overpower V *easily*. Conversely, if S is shorter than V and s/he indeed attempted to kill V, then there is only a *Very_Low* (VL) chance for S to overpower V *easily*.

C. Presumptions

The KB used in this work is assumed to possess the following properties.
1) *There are no cycles in the KB:* In other words, there are no self-referencing rules in the KB. This means that an antecedent variable cannot be its own consequence. This is required to support representation of causality in variable relations.
2) *Incomplete assignments of certainty degrees:* It is not required that every possible combination of antecedent and assumption values has to be assigned a certainty degree as the number of combinations will increase 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.

D. Fuzzy Complex Numbers

Within this work, generic MFs may involve both vague and uncertain information. In dynamically instantiating and composing those potentially relevant MFs into plausible-scenario descriptions, such inexact knowledge and data need to be combined and propagated throughout the emerging-model space. In order to achieve this, a novel framework of FCNs is proposed, with the significant capability of representing two types of uncertainty conjunctively. FCNs inherit from the real complex numbers.

Definition 4: An FCN, i.e., $\tilde{z}$, is defined in the form of

$$\tilde{z} = \tilde{a} + i\tilde{b} \qquad (2)$$

where both $\tilde{a}$ and $\tilde{b}$ are fuzzy numbers with membership functions $\mu_{\tilde{a}}(x)$ and $\mu_{\tilde{b}}(x)$, with regard to a given domain variable x. Term $\tilde{a}$ is the real part of $\tilde{z}$, while $\tilde{b}$ represents the imaginary part, i.e., $\text{Re}(\tilde{z}) = \tilde{a}$, and $\text{Im}(\tilde{z}) = \tilde{b}$.

Importantly, in general, for a given $\tilde{z}$, both $\text{Re}(\tilde{z})$ and $\text{Im}(\tilde{z})$ are fuzzy numbers. However, if $\text{Re}(\tilde{z})$ and $\text{Im}(\tilde{z})$ degenerate to real numbers, then $\tilde{z}$ degenerates to a real complex number. In addition, if $\tilde{b}$ does not exist, $\tilde{z}$ degenerates to a fuzzy number.

Further, if $\tilde{b}$ does not exist and $\tilde{a}$ itself degenerates to a real number, then $\tilde{z}$ degenerates to a real number.

In a given application, both the real and imaginary parts of an FCN can be assigned with their embedded semantic meaning. In particular, to support CM, the real part is utilized to represent the certainty degree of a certain piece of information, while the imaginary part represents the fuzzy-matching degree of a given piece of evidence or a derived piece of information using a particular rule instance of a certain MF. As such, FCNs offer an efficient common scheme to represent all four previously mentioned types of inexact information. Note, particularly, the following points.

1) *Vagueness*: In CM, the imaginary part only exists when fuzzy matching is performed between two pieces of information. 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 that represents this evidence is equal to 1. For example, given a fact, i.e., f_1: Bob is *tall*, this piece of information is represented as: $\tilde{z} = 1$. However, when a piece of evidence, i.e., e_1: victim is *very_tall*, is collected, f_1 matches e_1 with a fuzzy-matching degree $\tilde{b}$, such that if f_1 is activated, the FCN attached to f_1 will be written as $\tilde{z} = 1 + i\tilde{b}$.

2) *Uncertainty*: If a piece of information (e.g., a rule instance) 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.

3) *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 for the third type and is a fuzzy number for the fourth.

Notationwise, it may be the case that FCNs can be re-represented using a 2-D vector. However, the introduction of FCNs allows an effective extension of conventional complex number concepts and calculus in fuzzy terms. This is, in principle, independent of the development of fuzzy-CM techniques, thereby making a potential useful contribution to the general research on fuzzy sets and systems. The modulus of FCNs well preserves the physical interpretation of the conjunction of different information measures that may be gauged in dissimilar dimensions. Indeed, the framework of FCNs can be extended to represent arbitrary n-types of inexact information. In particular, in supporting (fuzzy) model composition (see Section IV-B), the modulus offers a global measure to rank the possible assignments. This helps to effectively and efficiently maintain the emerging scenario space, since only the most-likely scenario descriptions are necessarily generated in the first instance. In addition, the implementation of FCNs is simple and efficient, where a single system index is sufficient to assess each multidimensional FCN. With an alternative representation scheme like a simple vector, additional effort is needed to record and propagate dimensional-specific indices within the process of model composition. This would reduce the transparency and interpretability of the modeling process to users, in addition to an increase in computational cost.

In general, the framework of fuzzy CM works without the need of FCNs if there is only one type of inexact information to be considered. However, the notion of FCNs provides a more convenient and effective means of handling multimodal inexact information within the process of CM, especially with regard to the representation and propagation of such information.

E. Fuzzy-Complex-Number Operators and Their Properties

To support the work of CM, the basic operations on the proposed FCNs need to be developed. Note that the *rectangular FCNs*, which are proposed in [2], are represented in a form that is similar to the proposed work (but have different interpretations). In addition, when performing addition and subtraction on these two types of FCNs, the same results may be obtained. However, in multiplication and division, this is not the case. The *rectangular FCNs* are defined as type-1 fuzzy sets, and the basic arithmetic operations upon them are developed using the extension principle. Here, the operations on the proposed FCNs are a straightforward extension of those on real complex numbers.

Definition 5: Let $\tilde{z}_1 = \tilde{a} + i\tilde{b}$ and $\tilde{z}_2 = \tilde{c} + i\tilde{d}$ be two FCNs, where $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, and $\tilde{d}$ are fuzzy numbers with membership functions $\mu_{\tilde{a}}(x)$, $\mu_{\tilde{b}}(x)$, $\mu_{\tilde{c}}(x)$, and $\mu_{\tilde{d}}(x)$, respectively. The basic arithmetic operations on $\tilde{z}_1$ and $\tilde{z}_2$ are defined as follows.

1) *Addition:* It is defined as

$$\tilde{z}_1 + \tilde{z}_2 = (\tilde{a} + \tilde{c}) + i(\tilde{b} + \tilde{d}) \tag{3}$$

where $\tilde{a} + \tilde{c}$ and $\tilde{b} + \tilde{d}$ are derived fuzzy numbers with the following membership functions, respectively:

$$\mu_{\tilde{a}+\tilde{c}}(y) = \bigvee_{y=x_1+x_2} (\mu_{\tilde{a}}(x_1) \wedge \mu_{\tilde{c}}(x_2))$$

$$\mu_{\tilde{b}+\tilde{d}}(y) = \bigvee_{y=x_1+x_2} (\mu_{\tilde{b}}(x_1) \wedge \mu_{\tilde{d}}(x_2)). \tag{4}$$

2) *Subtraction:* It is defined as

$$\tilde{z}_1 - \tilde{z}_2 = (\tilde{a} - \tilde{c}) + i(\tilde{b} - \tilde{d}) \tag{5}$$

where $\tilde{a} - \tilde{c}$ and $\tilde{b} - \tilde{d}$ are derived fuzzy numbers with the following membership functions, respectively:

$$\mu_{\tilde{a}-\tilde{c}}(y) = \bigvee_{y=x_1-x_2} (\mu_{\tilde{a}}(x_1) \wedge \mu_{\tilde{c}}(x_2))$$

$$\mu_{\tilde{b}-\tilde{d}}(y) = \bigvee_{y=x_1-x_2} (\mu_{\tilde{b}}(x_1) \wedge \mu_{\tilde{d}}(x_2)). \tag{6}$$

3) *Multiplication:* It is defined as

$$\tilde{z}_1 \times \tilde{z}_2 = (\tilde{a}\tilde{c} - \tilde{b}\tilde{d}) + i(\tilde{b}\tilde{c} + \tilde{a}\tilde{d}) \tag{7}$$

where $\tilde{a}\tilde{c} - \tilde{b}\tilde{d}$ and $\tilde{b}\tilde{c} + \tilde{a}\tilde{d}$ are derived fuzzy numbers with the following membership functions, respectively:

$$\mu_{\tilde{a}\tilde{c}-\tilde{b}\tilde{d}}(y) = \bigvee_{y=x_1x_2-x_3x_4} (\mu_{\tilde{a}}(x_1) \wedge \mu_{\tilde{c}}(x_2)$$

$$\wedge \mu_{\tilde{b}}(x_3) \wedge \mu_{\tilde{d}}(x_4))$$

$$\mu_{\tilde{b}\tilde{c}+\tilde{a}\tilde{d}}(y) = \bigvee_{y = x_1 x_2 + x_3 x_4} (\mu_{\tilde{b}}(x_1) \wedge \mu_{\tilde{c}}(x_2)$$
$$\wedge \mu_{\tilde{a}}(x_3) \wedge \mu_{\tilde{d}}(x_4)). \tag{8}$$

4) *Division:* It is defined as

$$\frac{\tilde{z}_1}{\tilde{z}_2} = \left(\frac{\tilde{a}\tilde{c} + \tilde{b}\tilde{d}}{\tilde{c}^2 + \tilde{d}^2} \right) + i \left(\frac{\tilde{b}\tilde{c} - \tilde{a}\tilde{d}}{\tilde{c}^2 + \tilde{d}^2} \right). \tag{9}$$

For notational simplicity, let $\tilde{t}_1 = (\tilde{a}\tilde{c} + \tilde{b}\tilde{d})/(\tilde{c}^2 + \tilde{d}^2)$, and $\tilde{t}_2 = (\tilde{b}\tilde{c} - \tilde{a}\tilde{d})/(\tilde{c}^2 + \tilde{d}^2)$, where $\tilde{t}_1$ and $\tilde{t}_2$ are derived fuzzy numbers with the following membership functions, respectively:

$$\mu_{\tilde{t}_1}(y) = \bigvee_{y = \frac{x_1 x_3 + x_2 x_4}{x_3^2 + x_4^2}, x_3^2 + x_4^2 \neq 0} (\mu_{\tilde{a}}(x_1) \wedge \mu_{\tilde{b}}(x_2)$$
$$\wedge \mu_{\tilde{c}}(x_3) \wedge \mu_{\tilde{d}}(x_4))$$

$$\mu_{\tilde{t}_2}(y) = \bigvee_{y = \frac{x_2 x_3 - x_1 x_4}{x_3^2 + x_4^2}, x_3^2 + x_4^2 \neq 0} (\mu_{\tilde{a}}(x_1) \wedge \mu_{\tilde{b}}(x_2)$$
$$\wedge \mu_{\tilde{c}}(x_3) \wedge \mu_{\tilde{d}}(x_4)). \tag{10}$$

5) *Modulus:* Given $\tilde{z} = \tilde{a} + i\tilde{b}$, the modulus of $\tilde{z}$ is defined to be

$$|\tilde{z}| = \sqrt{\tilde{a}^2 + \tilde{b}^2} \tag{11}$$

with the following membership function:

$$\mu_{|\tilde{z}|}(y) = \bigvee_{y = \sqrt{x_1^2 + x_2^2}} (\mu_{\tilde{a}}(x_1) \wedge \mu_{\tilde{b}}(x_2)). \tag{12}$$

Let $\tilde{z}_1 = \tilde{a} + i\tilde{b}$, $\tilde{z}_2 = \tilde{c} + i\tilde{d}$, and $\tilde{z}_3 = \tilde{e} + i\tilde{f}$ be three FCNs, where $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}, \tilde{e}$, and $\tilde{f}$ are fuzzy numbers with membership functions $\mu_{\tilde{a}}(x), \mu_{\tilde{b}}(x), \mu_{\tilde{c}}(x), \mu_{\tilde{d}}(x), \mu_{\tilde{e}}(x)$, and $\mu_{\tilde{f}}(x)$, respectively. The algebraic properties of the proposed FCNs are summarized as follows, and due to space limitations, the proofs for these theorems are omitted.

Theorem 1: If $\tilde{z}_1 = \tilde{a} + i\tilde{b}$, and $\tilde{z}_2 = \tilde{c} + i\tilde{d}$ are FCNs, then so are $\tilde{z}_1 + \tilde{z}_2$, $\tilde{z}_1 - \tilde{z}_2$, $\tilde{z}_1\tilde{z}_2$, and $|\tilde{z}|$.

Theorem 2: Associativity, i.e., $\tilde{z}_1 * (\tilde{z}_2 * \tilde{z}_3) = (\tilde{z}_1 * \tilde{z}_2) * \tilde{z}_3$, holds if $* = +$.

Theorem 3: Commutativity, i.e., $\tilde{z}_1 * \tilde{z}_2 = \tilde{z}_2 * \tilde{z}_1$, holds for $* \in \{+, \times\}$.

Theorem 4: Distributivity, i.e., given $\tilde{a} > 0$ or $\tilde{a} < 0$ and $\tilde{b} > 0$ or $\tilde{b} < 0$, when $\tilde{c}$ and $\tilde{e}$ have the same sign (they are both either a positive or negative fuzzy number), and $\tilde{d}$ and $\tilde{f}$ also have the same sign, then $\tilde{z}_1 \times (\tilde{z}_2 + \tilde{z}_3) = \tilde{z}_1 \times \tilde{z}_2 + \tilde{z}_1 \times \tilde{z}_3$.

IV. FUZZY COMPOSITIONAL MODELING

The main task of a *fuzzy compositional modeler* is to automatically generate different scenarios that can explain the available evidence. The overall architecture of such a modeler is shown in Fig. 3. The contents of the KB have been introduced previously. Given a new or ongoing investigation in which an initial set of evidence E has been collected by the investigators, the *fuzzy-MF retrieval* component initially retrieves those most-relevant MFs
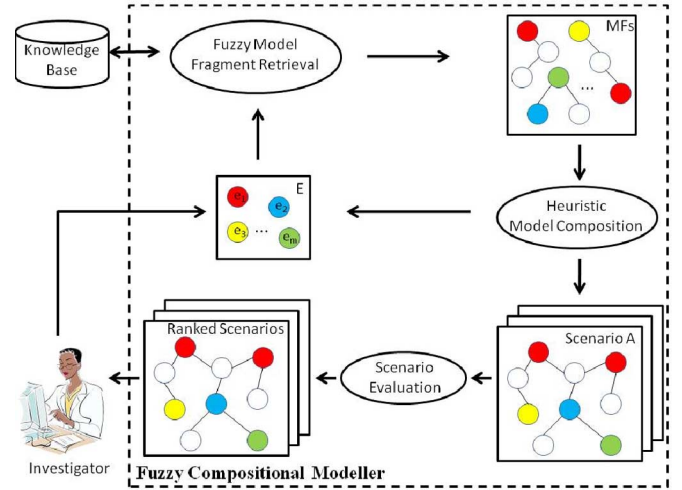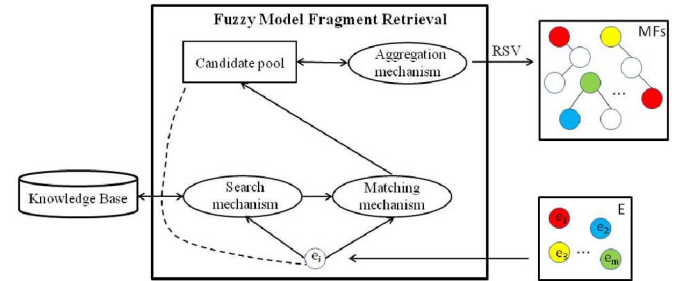

Fig. 3. Outline of the FCM approach.

Fig. 4. Fuzzy-MF retrieval.

from the KB. Once the relevant MFs have been retrieved, the *heuristic-model-composition* component instantiates and synthesizes these MFs into a large scenario space that contains a range of plausible scenarios. The generated scenario space describes how plausible pieces of evidence and newly derived information may casually be related to one another. It is therefore practically important to incorporate a means of evaluation for such generated scenarios so that the generated information remains manageable by the investigators. Finally, ranked plausible scenarios are fed back to the investigators for further analysis or to determine future investigating actions. Of course, further collected evidence can be added to E to start a new synthesis cycle. Technical details to implement the underlying inference mechanisms of FCM are described in the following sections.

A. Fuzzy-Model Fragments Retrieval

Given a set of collected evidence E and a KB, FCM begins by retrieving from the KB those MFs that are most likely to be relevant to the available evidence. However, due to the involvement of inexact information, a precise and certain match between the available evidence and the elements in the KB cannot, in general, be expected. Partial matching may be all that is feasible and may suffice for many application problems. The process of retrieving partially matched and, hence, partly instantiated MFs is outlined in Fig. 4.

The collected evidence (E) is supplied to the system in a piecewise fashion. Given the KB, the *search component* identifies those MFs that are related to the involved concepts/object in the given piece of evidence (e_i). Then, the selected MFs and e_i, which led to such selection, are both fed to the *matching component* to compute their corresponding degrees of match. These retrieved MFs together with the resulting matching degrees are stored in a candidate pool for creation of the possible scenarios, along with other candidates using other pieces of evidence. After this, the *aggregation component* is employed to calculate an overall relevance degree of each candidate MF to all the given evidence, which is termed the retrieval status value (RSV) for easy reference. Only those candidate MFs whose derived RSVs are greater than a threshold will be passed to the next step. Finally, the output of this entire iterative process is an ordered set of retrieved MFs, which have been at least partially instantiated in relation to E.

1) Sematic Matching: In this work, a predefined taxonomy T (see Section III-A) is employed to perform sematic matching between two concepts/objects. Any two nodes within T may be semantically considered similar to some extent if they are linked, namely, one being another's antecedent or descendant. Without losing generality, best-first search is adapted to identify the position of the required concept in T. The search starts from the root node of T. At a common level of the taxonomical tree, the higher the weight a child node has with respect to its parent, the earlier it is checked to see if it is of semantic similarity with the concept being searched for. Note that, in general, to avoid practical difficulties in acquiring the required taxonomy, it is allowed to have similar or even the same linguistic term appearing in different parts of T.

Definition 6: Given two nodes c and c' of a weighted taxonomy T, the semantic similarity between c and c' is computed by

$$S_v(c, c') = \sum_{i \in N_{cc'}} \frac{w_i}{|N_{cc'}|} \qquad (13)$$

where $N_{cc'}$ is the collection of all the edges of the shortest path connecting c and c' in T, w_i is the weight on edge i, and $|N_{cc'}|$ is the cardinality of $N_{cc'}$. $S_v(c, c') = 0$, if there is no generational link between c and c'. $S_v(c, c') = 1$, if $c = c'$.

Obviously, $S_v(c, c') \in [0, 1]$. In particular, $S_v(c, c') = 1$, if $c \neq c'$, and every edge has a weight of 1. This definition captures intuition well. Let us consider Fig. 2, for example, given two nodes "Cola" and "Liquid," their semantic similarity can be computed such that

$$S_v(\text{Cola}, \text{Liquid}) = \frac{0.8 + 0.76}{2} = 0.78.$$

2) Fuzzy-Set Matching: As aforementioned, variables involved in fuzzy MFs are allowed to be boolean, fuzzy, or mixed. Boolean variables can only be evaluated to be either true or false, whereas the value of a fuzzy variable may be represented by a fuzzy set. Given a single piece of evidence, x is v, the semantic match between x and antecedent/concequent variables in an MF is performed as outlined above. Fuzzy-set matching addressed here calculates the similarity degree of variable values. If x is

a boolean variable, the value of x must exactly match the value of the antecedent/consequent variable, and the matching degree between them is therefore either 1 or 0. For a fuzzy variable, the degree of matching (of two fuzzy sets) is a value in the range $[0, 1]$. In this work, the *Hausdorff* distance [15] is employed to measure the fuzzy-set matching degrees.

Definition 7: Given two triangular fuzzy sets $A = [a_1, a_2, a_3]$, and $B = [b_1, b_2, b_3]$, $A \neq B$, the *Hausdorff* distance between them is defined as

$$S_d(A, B) = \max\{d(A, B), d(B, A)\}$$
$$= \max\left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\} \qquad (14)$$

where $d(a, b)$ is the normalized absolute distance between parameters a and b, i.e.,

$$d(a, b) = \frac{|a - b|}{\max\{|b_3 - a_1|, |a_3 - b_1|\}}.$$

Definition 8: Given two fuzzy sets A and B, the matching degree, i.e., S_f, between them is given by

$$S_f(A, B) = 1 - S_d(A, B). \qquad (15)$$

Since each MF may employ a set of rule instances, given one piece of evidence e, more than one rule instance may be instantiated with different matching degrees. In this work, the largest fuzzy-set matching degree will be taken to represent the overall S_f between the MF and e.

Based on the above semantic and fuzzy-set matching, the overall matching degree between "x is v" and "x' is v'" is deemed to be aggregated by using the algebraic *product* operator

$$S(x : v, x' : v') = S_v(x, x') \oplus S_f(x : v, x' : v'). \qquad (16)$$

Here, the *product* operator is adopted for aggregation because boolean predicates may be involved, such that the S_f of boolean predicates is either 1 or 0. It is obvious that if $S_f = 1$, then the use of other aggregation operators, such as *addition* or *max,* may lead to the S_f of boolean predicates dominating the aggregation process. In addition, if $S_f = 0$, the overall matching degree will become 0 by using the *product* operator. This clearly reflects the intuition well.

3) Aggregation: The above steps are iteratively carried out until all elements in E have been examined. For each $e_i \in E$, a set of relevant MFs may be instantiated and returned. Each such MF will have a relevance degree attached with respect to e_i. However, the antecedence of an MF often involves more than one atomic predicate that are connected by conjunctive/disjunctive operators. Hence, those relevance degrees associated with the individual antecedent predicates need to be aggregated to derive the final RSV and stored with the MF in the set of retrieved MFs in preparation for composition.

The individual relevance degrees of the n atomic predicates in one MF, i.e., $S_1, S_2, \ldots, S_n$, are herein aggregated by

$$f_i = f_{i-1} + S_i - f_{i-1} \times S_i \qquad (17)$$

where $f_0 = S_0 = 0$, and $i = 1, 2, \ldots, n$, with n being the number of activated atomic predicates in an MF and f_n being the

final RSV of that MF. This aggregation scheme once again reflects the intuition that the more evidence there exists to support an MF, the higher relevance that MF is to be involved in the scenario to build.

B. Model Composition

Once the potentially relevant MFs have been retrieved and at least partially instantiated, a mechanism is required to dynamically synthesize those MFs into consistent plausible scenarios. This section presents such a model-composition technique. First, the inference mechanism to generate plausible scenarios is described. Next, several filtering methods are introduced to reduce the solution space. In particular, during the process of composition, vague and uncertain knowledge and data need to be propagated from individually instantiated MFs to their related ones. Hence, this section also describes how inexact information captured in FCNs may be combined and propagated through constraints.

1) Creation of Model Space: Composition of a scenario space, or model space, in general, requires a means for the scenario elements to interact. This is facilitated by the use of the shared variables. The retrieved MFs must have shared variables with the collected set of evidence and, normally, have shared variables among themselves (if they are relevant in describing a common scenario). In CM, a node in the space of plausible-scenario descriptions (which is called emerging scenario space, hereafter) represents a proposition, and each has an FCN attached. After MF retrieval, those nodes that share common variables with the available evidence will have initial FCNs attached, thus indicating the vague and uncertain information involved. These initial FCNs are the starting point of model composition, and they will be spread gradually throughout the scenario generation.

Given an MF, if initial FCNs are associated with the values of antecedent variables, then the consequent variable's value can be derived by using the compositional rule of inference [34]. This process is named forward propagation of FCNs. Conversely, if the initial FCN is associated with the value of the consequent variable, then a backward propagation of FCNs is needed to derive the values of the antecedent variables. However, due to the lack of inverse operators over fuzzy sets, there is no general method to compute the exact values of antecedent variables directly. This is especially the case when there are more than one antecedent variable involved.

To address this problem, fuzzy constraints embedded within MFs are not used to derive the unknown values of the related variables. Instead, such constraints over antecedent and consequent variables are causally employed to check for consistency among their respective possible values. Once the quantity space of a variable within a retrieved MF has been revised and some spurious values removed, such changes are propagated to those MFs that share the same variables. In this work, the CM process is completed by iteratively passing the revised MFs (actually, the revised quantity space of the variables involved in these MFs) through the KB. Note that such iterative passages of reduced quantity spaces are not limited to the current set of retrieved

---

Algorithm 1 Waltz()

1: $activeCon \leftarrow$ a queue of retrieved MFs
2: **while** $activeCon \neq \emptyset$ **do**
3:   remove constraint C from $activeCon$
4:   $CHANGED \leftarrow Revise(C)$
5:   **for all** v_i in $CHANGED$ **do**
6:     **for all** $C' \neq C$ which involves v_i **do**
7:       **if** $C' \notin activeCon$ **then**
8:         add C' to $activeCon$
9:       **end if**
10:     **end for**
11:   **end for**
12: **end while**

Algorithm 2 Revise($C(v_1, \ldots, v_k)$)

1: $CHANGED \leftarrow \emptyset$
2: **for all** initial pieces of evidence $e(v_e)$ in E **do**
3:   **if** $v_e \cap \{v_1 \ldots v_k\} \neq \emptyset$ and $S_c = \emptyset$ **then**
4:     $S_c \leftarrow Refine(C, e)$
5:     add $v_1, \ldots, v_k$ to $CHANGED$
6:   **end if**
7: **end for**
8: **for all** constraints $C_i(v_{i1}, \ldots, v_{in}) \neq C(v_1, \ldots, v_k)$ in $revisedCon$ and $\{v_{i1}, \ldots, v_{in}\} \cap \{v_1, \ldots, v_k\} \neq \emptyset$ **do**
9:   $tmp\_S_c = Refine(C, C_i)$
10:   **if** $tmp\_S_c = \emptyset$ **then**
11:     halt{No solution}
12:   **else if** $S_c \neq tmp\_S_c$ **then**
13:     $S_c = tmp\_S_c$
14:     add $v_1, \ldots, v_k$ to $CHANGED$
15:   **end if**
16: **end for**
17: add C to $revisedCon$
18: **return** $CHANGED$

MFs, but the entire KB to enable further search of other MFs, that are not instantiated by the present evidence. This process continues until no further changes are produced.

To implement the above process efficiently, Waltz algorithm [30] (as shown in Algorithm 1) is applied. Running the Waltz algorithm equivalently adds or removes instantiated MFs dynamically. The addition of new MFs is more likely as the propagation of variable values may instantiate other variables that have not been activated so far. However, the removal of certain MFs may also take place if a certain variable's quantity space is reduced to empty, which means that the constraints given so far are inconsistent. At termination of the algorithm, a set of consistent and the most-likely scenario descriptions that explain the given evidence is generated. Note that the constraints are not used to find unknown values but employed to work as filters to check for consistency among given values. The actual filtering mechanism is explained below.

2) Filtering Techniques: The main component of Waltz algorithm is that it applies the *Revise* method (see Algorithm 2) to each retrieved MF. As aforementioned, each MF may have several rule instances or candidate assignments attached. Given an MF, if it is activated by the collected evidence, each candidate assignment within this MF will go through the filters (see
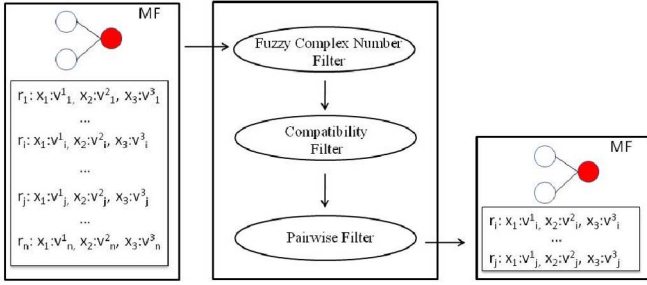
Fig. 5.   Filtering the candidate assignments.

below) and the output of the *Refine* method is termed survival assignments and saved in a buffer S_c. Note that if an MF has been revised, there is no need to check all candidate assignments from scratch, but only those in S_c need to be further checked by other filters.

As illustrated in Fig. 5, there are three filters that candidate assignments need to pass, namely, *FCN filter*, *compatibility filter*, and *pairwise filter*. *FCN filter* is employed to check the consistency between the FCNs associated with individual variables within one MF. Any candidate assignment involving an inconsistent FCN is ruled out. Additionally, the remaining assignments are ranked with respect to the moduli of their FCNs. In doing so, the most-likely scenario descriptions can be generated first, and the generation of alternative plausible scenarios is postponed until more likely ones have to be discarded.

In any knowledge-based approach, inconsistent information may lead to not only unexpected and unsatisfactory results but to wasted computational resources as well. When inconsistency is detected, such information is recorded and utilized to impose further restriction over future candidate assignments. This is implemented as the *compatibility filter* by introducing a specific type of MF, which is called *nogood* MF in the KB, which collects all detected inconsistencies. Passing candidate assignments through this filter ensures that each surviving assignment is consistent with the knowledge accumulated so far. For example, the following *nogood* MF shows that it is inconsistent for V to commit suicide and to be the victim of homicide by M at the same time.

IF {commits suicide(V), homicide by(V,M)}
THEN {nogood}
Distribution nogood {r_1: true, true → true: 1}.

Finally, in the emerging scenario space, if there are two or more surviving MFs sharing a common variable, the assignments of that variable in those MFs must be identical to each other. This constraint is checked by the *pairwise filter*. After this, the global consistency of the emerging scenario space is guaranteed (due to the *compatibility filter* and *pairwise filter*).

In terms of algorithmic complexity, suppose that there are n variables and m MFs in the KB and that in the worst case, each of the m MF consists of k plausible rule instances. In the worse case, the algorithm may use $\mathbf{O}(m^2 s_1 + mkns_2)$ space, where s_1 and s_2 are the unit memory space for an MF and a scenario node, respectively. In addition, the time complexity to execute this algorithm is $(m + n(m-1)) \times k \times T \approx \mathbf{O}(mnkT)$. Due

to space constraints, the proofs for these results are omitted here. In essence, due to the Waltz's algorithm, the worst-case complexity of the proposed model-composition process is linear with regard to the number of variables for both time required and space used.

3) Propagation of Fuzzy Complex Numbers: In fuzzy CM, the vague and uncertain information (which is concisely represented in terms of FCNs) needs to be combined and propagated through relations embedded in MFs. Within an emerging scenario space, each node represents a proposition with an FCN attached. Each rule instance, in any generic MF, is also associated with an FCN, which is denoted by $\tilde{z}_r$, indicating the certainty degree of the corresponding causal proposition. Thus, the model-composition process combines the FCNs attached to the (antecedents or consequent) variables and the FCN attached to each instantiated rule instance ($\tilde{z}_r$).

For simplicity, the following rule instance in an MF:

$$\text{IF } x_1 \text{ is } A_{j_1}^1 \oplus \cdots \oplus x_n \text{ is } A_{j_n}^n, \text{ THEN } y \text{ is } B_{j_{n+1}} (\tilde{z}_r)$$

is rewritten as

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

where $p_1, \ldots, p_n$ are the antecedent propositions, each of which has an attached FCN, $\tilde{z}_{p_j} = \text{Re}(\tilde{z}_{p_j}) + i\text{Im}(\tilde{z}_{p_j}), j \in \{1, 2, \ldots, n\}$, c is the consequent proposition, and $\tilde{z}_r$ is the FCN attached to the rule instance. As stated previously, $\oplus$ in the antecedence may be interpreted as either a conjunctive or disjunctive operator.

Definition 9: If $\oplus$ is a conjunctive operator, then the aggregated FCN of the antecedent is given by

$$\tilde{z}_{\text{antecedent}} = \min(\text{Re}(\tilde{z}_{p_1}), \ldots, \text{Re}(\tilde{z}_{p_n}))$$
$$+ i \min(\text{Im}(\tilde{z}_{p_1}), \ldots, \text{Im}(\tilde{z}_{p_n})). \quad (18)$$

If $\oplus$ is a disjunctive operator, then the aggregated FCN of the antecedent is given by

$$\tilde{z}_{\text{antecedent}} = \max(\text{Re}(\tilde{z}_{p_1}), \ldots, \text{Re}(\tilde{z}_{p_n}))$$
$$+ i \max(\text{Im}(\tilde{z}_{p_1}), \ldots, \text{Im}(\tilde{z}_{p_n})). \quad (19)$$

Since a rule instance in an MF only has the certainty degree attached, i.e., $\tilde{z}_r = \text{Re}(\tilde{z}_r)$, the newly derived FCN attached to the consequence is deemed to be

$$\tilde{z}_{\text{new}} = \tilde{z}_{\text{antecedent}} \times \tilde{z}_r$$
$$= \text{Re}(\tilde{z}_{\text{antecedent}}) \times \tilde{z}_r + i \text{Im}(\tilde{z}_{\text{antecedent}}) \times \tilde{z}_r \quad (20)$$

where $\text{Re}(\tilde{z}_{\text{antecedent}})$ and $\text{Im}(\tilde{z}_{\text{antecedent}})$ are fuzzy numbers, in general.

There are two types of propagation that need to be discussed separately: propagating an FCN from antecedent to consequent and the reverse.

BackwardPropagation: This procedure is employed to induce the domain variables and their states, which might have led to the available evidence that matches the consequent of an MF. Plausible causes can be identified by instantiating the conditions and assumptions of the MF. As indicated earlier, given $\tilde{z}_c$ and $\tilde{z}_r$, no general method exists to derive

---

Algorithm 3 BackwardPropagation($\tilde{z}_c, \tilde{z}_r$)

1: $Re(\tilde{z}_{antecedent}) \leftarrow \emptyset$
2: $Im(\tilde{z}_{antecedent}) \leftarrow \emptyset$
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:         $Re(\tilde{z}_{antecedent}) = V_{FN_j}$
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:       $Im(\tilde{z}_{antecedent}) = V_{FN_j}$
13:    **end if**
14: **end for**
15: **return** $\tilde{z}_{antecedent}$

Algorithm 4 ForwardPropagation($\tilde{z}_{p_1}, \ldots, \tilde{z}_{p_n}, \tilde{z}_r$)

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

---

the exact value of $\tilde{z}_{\mathrm{antecedent}}$ in a closed form. In doing so, for computational simplicity, it is assumed that both the real and the imaginary part of $\tilde{z}_{\mathrm{antecedent}}$ take values from a certain fixed quantity space: $Q_{FN} = \{V_{FN_1}, \ldots, V_{FN_n}\}$, where $V_{FN_j}, j \in \{1, 2, \ldots, n\}$ may be specified with respect to a given problem. Of course, in theory, it is not necessary for the real and the imaginary part to take values from a common quantity space.

Algorithm 3 is constructed by ensuring the generality that each element in Q_{FN} may be the possible value of $\tilde{z}_{\mathrm{antecedent}}$. Thus, all such values are checked using (20) as a constraint. By applying (20), a derived fuzzy number does not necessarily belong to Q_{FN}. The element that best satisfies this constraint (i.e., with the largest matching degree) when given $\tilde{z}_c$, is selected to represent $\tilde{z}_{\mathrm{antecedent}}$. Note that $Re(\tilde{z}_c)$ and $Im(\tilde{z}_c)$ are checked against the constraint separately.

Once $\tilde{z}_{\mathrm{antecedent}}$ is obtained, the next step is to derive the individual FCNs associated with each antecedent variable. Without losing generality, taking the conjunctive operator as an example, the following constraints are introduced:

$$\min(Re(\tilde{z}_{p_1}), \ldots, Re(\tilde{z}_{p_n})) = Re(\tilde{z}_{\mathrm{antecedent}})$$
$$\min(Im(\tilde{z}_{p_1}), \ldots, Im(\tilde{z}_{p_n})) = Im(\tilde{z}_{\mathrm{antecedent}}). \quad (21)$$

These constraints state that, for any $(V_{FN_m}, \ldots, V_{FN_k}) \in Q_{FN} \times \cdots \times Q_{FN}$, if $\min(V_{FN_m}, \ldots, V_{FN_k}) = Re(\tilde{z}_{\mathrm{antecedent}})$, then $V_{FN_m}, \ldots, V_{FN_k}$ are possible solutions for the real parts of $\tilde{z}_{p_1}, \ldots, \tilde{z}_{p_n}$. Similarly, the possible solutions of the imaginary parts can be obtained. Note that, if $\oplus$ is interpreted as a disjunctive operator, all that is needed is to change the *min* in (21) to *max*.

ForwardPropagation: This procedure applies logical deduction to all those MFs in the KB whose conditions match the existing nodes in the emerging scenario space. As described in Algorithm 4, $\tilde{z}_{\mathrm{antecedent}}$ and $\tilde{z}_r$ are known, and the FCN of the consequent variable, i.e., $\tilde{z}_{\mathrm{new}}$, is computed by applying (20). Given the fixed Q_{FN}, once $\tilde{z}_{\mathrm{new}}$ is computed, it is used to choose those elements of Q_{FN}, such that the selected ones will receive the highest matching degree with $\tilde{z}_{\mathrm{new}}$. Again, the real and imaginary parts are checked separately when implemented.

4) Fuzzy-Complex Numbers Updating: In generating the scenario space, if instantiating one MF leads to a node that

has already existed (i.e., having been activated with another MF), such that two instantiated MFs share a common variable value, then the existing FCNs associated with this node needs to be updated for consistency. Suppose that for a certain variable to take value A, and that the following two FCNs are obtained through different inference procedures

$$\tilde{z}_A = V_{FN_i} + iV_{FN_j}, \tilde{z}'_A = V_{FN_p} + iV_{FN_q}$$

where $V_{FN_i}, V_{FN_j}, V_{FN_p}$, and V_{FN_q} denote different elements in the predefined Q_{FN}, respectively. Given that the modulus of an FCN indicates the overall confidence in the information content of its associated variable value. Thus, 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 $\tilde{z}'_A$ will be kept as possible FCNs. However, if $|\tilde{z}_A| \neq |\tilde{z}'_A|$, then the updated FCN of this variable taking value A is obtained by

$$\tilde{z}''_A = \min(V_{FN_i}, V_{FN_p}) + i\min(V_{FN_j}, V_{FN_q}). \quad (22)$$

This has an intuitive appeal because if (22) holds, then both $\tilde{z}_A$ and $\tilde{z}'_A$ will also hold.

V. APPLICATION TO CRIME INVESTIGATION

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 or prevention [27]. However, humans are relatively inefficient at hypothetical reasoning, especially for cases where there are a large amount of available data and knowledge involved. Intelligence experts have commented that failure to detect plausible threats is not so much due to a lack of data but more due to difficulties in relating to and reasoning about the available data [1], [8]. There are many potential sources of variability in precision, including vaguely defined concepts, quantities and specifications of importance, and certainty. It is, therefore, very interesting to examine if the proposed fuzzy compositional modeler can be employed to assist investigators in generating plausible-scenario descriptions and analyzing them objectively. Note that this work has been implemented computationally in Java. Because of space constraints, only a proof-of-concept example is provided here for

TABLE I
FUZZY VARIABLES

| Vars	Type	Interpretation	Domain
x_1	Fuzzy	Degree of Z shattered	{very_low, low, medium, quite_high, high}
x_2	Fuzzy	Sound volume of the explosion	{low, medium, loud}
x_3	Fuzzy	Distance between P and Z	{near, medium, far}
x_4	Fuzzy	Power level of the explosion	{low, medium, high}
x_5	Fuzzy	The amount of X	{none, few, several, many, a_lot}
x_6	Fuzzy	The amount of Y	{none, few, several, many, a_lot}
x_7	Boolean	X is a liquid	{true, false}
x_8	Boolean	Y is a hydrogen_peroxide	{true, false}
x_9	Boolean	Mix of X and Y	{true, false}
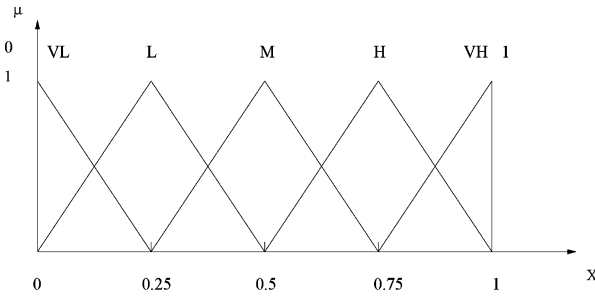
Fig. 6.   Fuzzy quantity space for both certainty and fuzzy-matching degrees.

illustration purposes. Details of a scaled-up prototype system is beyond the scope of this paper.

A. Crime Scene

This illustrative example is extracted and adapted from a realistic case described in [12]. The case description is summarized as follows: *An explosion took place at an airport. Initially, the police found that the degree of shattering of a nearby window was quite_high. With further investigation, another two pieces of evidence were gathered. A suspect named Dave was trying to bring a bottle of cola on board an aircraft and was intercepted by the security staff. Also, a mobile phone was found on him. A few hours later, the security control centre reported that a small bag of hair dye was also found in a suitcase under the name of Dave.*

B. Knowledge Representation

The given pieces of evidence, $e_i, i = 1, 2, 3$, and some key MFs used in this example are detailed in Appendix A. In order to model this case, nine variables are extracted from the description, as listed in Table I. In addition, for illustrative simplicity, the fuzzy numbers used to form FCNs (namely, both certainty and fuzzy-matching degrees) are defined in the following $Q_{FN} = \{0, VL, L, M, H, VH, 1\}$ and shown in Fig. 6.

C. Fuzzy-Model Fragments Retrieval

Each element in E will go through the process of fuzzy MF retrieval, as previously described in Fig. 4. For simplicity, the

taxonomy of Fig. 2 is used in this example to illustrate semantic matching. Matching the first piece of evidence, i.e., e_1 (*degree_shattering(window)* is *quite_high*), against the KB leads to MF_2 being instantiated. In fact, the variable involved in e_1 and the consequent variable (x_1) of MF_2 are identical. Thus, by default, the semantic matching degree between e_1 and MF_2 is $S_v(v_{e_1}, x_1) = 1$.

The fuzzy-set matching degree between e_1 and the proposition "*degree_shattering(window)* is *high*" in MF_2 is obtained by using (14) and (15) is $S_f(v_{e_1} : \textit{quite\_high}, x_1 : \textit{high}) = 1 - 0.25 = 0.75$, where 0.25 is the *Hausdorff* distance between the corresponding two fuzzy sets. Note that e_1 matches all the rule instances in MF_2 with the same fuzzy-set matching degree (i.e., 0.75). Thus, the overall matching degree between the value of e_1 and that of the variable x_1 of MF_2 can be aggregated using (16), thus resulting in $S(e_1, x_1) = 1 \times 0.75 = 0.75$.

Sequentially, given e_2 (amount(cola) = *a_lot*), the antecedent variables x_5 and x_7 of MF_4 are instantiated. Matching them against the taxonomy of Fig. 2, the following semantic matching degrees are obtained, respectively: $S_v(v_{e_2}, x_5) = 1$ and $S_v(v_{e_2}, x_7) = 0.78$.

As aforementioned in Section IV-A2, given one piece of evidence e, if there are more than one rule instance of a certain MF being fired, the greatest fuzzy-set matching degree for each variable is intuitively taken to represent the overall S_f for that variable between the MF and e. Here, the greatest fuzzy-set matching degrees between e_2 and the rule instances of MF_4 with respect to different variables are obtained using (14) and (15), respectively, such that

$$S_f(v_{e_2} : a\_lot, x_5 : many) = 0.75$$
$$S_f(v_{e_2} : true, x_7 : true) = 1.$$

The overall matching degree between e_2 and x_5, and that between e_2 and x_7 of MF_4, are obtained by aggregation, which results in

$$S(e_2, x_5) = S_v(v_{e_2}, x_5) \oplus S_f(v_{e_2} : a\_lot, x_5 : many)$$
$$= 1 \times 0.75 = 0.75$$
$$S(e_2, x_7) = S_v(v_{e_2}, x_7) \oplus S_f(v_{e_2} : true, x_7 : true)$$
$$= 0.78 \times 1 = 0.78.$$

TABLE II
RESULTS OF BACKWARD PROPAGATION DUE TO e_1

Derived by Rule	Instantiated Antecedent		$Re(\tilde{z}_{antecedent})$	$Im(\tilde{z}_{antecedent})$
r_1	$x_2 : loud$	$x_3 : near$	1	H
r_2	$x_2 : loud$	$x_3 : medium$	1	1
r_3	$x_2 : loud$	$x_3 : far$	N/A	N/A
r_4	$x_2 : medium$	$x_3 : near$	1	1
r_5	$x_2 : medium$	$x_3 : medium$	N/A	1
r_6	$x_2 : medium$	$x_3 : far$	N/A	N/A
r_7	$x_2 : low$	$x_3 : near$	N/A	1
r_8	$x_2 : low$	$x_3 : medium$	N/A	N/A
r_9	$x_2 : low$	$x_3 : far$	N/A	N/A

Similarly, given e_3, MF$_4$ is fired again, and the overall matching degrees between e_3 and x_6, x_8 of MF$_4$ are calculated, respectively, as $S(e_3, x_6) = 1$ and $S(e_3, x_8) = 0.74$.

At termination of the above steps, all elements in E have been examined. As a result, two MFs (i.e., MF$_2$ and MF$_4$) are retrieved. Each MF can be instantiated by more than one piece of evidence; those individual matching degrees associated with different predicates of an MF are aggregated to derive its final RSV. For MF$_2$, only one consequent predicate is fired by e_1, and hence, there is no aggregation required.Hence, RSV$_{MF_2} = S(e_1, x_1) = 0.75$. For MF$_4$, four antecedent variables, namely, x_5, x_6, x_7, and x_8, are activated by e_2 and e_3. The individual matching degrees between these instantiated variables and the given evidence are $0.75, 0.78, 1$, and 0.74, respectively. These are aggregated by applying (17) to derive the overall RSVs of MF$_4$

$$\text{RSV}_{MF_4} = S(e_2, x_5) \oplus S(e_2, x_7) \oplus S(e_3, x_6) \oplus S(e_3, x_8) = 1.$$

Note that this RSV indicates the relevance degree of an MF to be involved in the scenario to be built. Only those MFs whose RSVs are greater than a threshold are kept for composition.

D. Generation of Plausible-Scenario Space

Because of the nature of evidence-driven scenario generation, backward propagation is performed first. Therefore, although MF$_2$ receives relatively lower RSV, it is still first instantiated. Matching the derived matching degree (i.e., 0.75) between e_1 and MF$_2$ back to the Q_{FN} leads to the identification of H being returned to the fuzzy-matching degree. Hence, the initial scenario space is first created with a node of "*degree_shattering*(*window*) is *high*," and it is associated with the initial FCN: $VH + iH$. Given this and the instantiated MF$_2$, backward propagation is performed from x_1. This leads to x_2 and x_3 being added to the emerging scenario space. By applying Algorithm 3 to MF$_2$, the results of Table II are obtained.

In order to avoid generating unnecessary explanations, the modeler produces only the current best or the most-plausible explanations in the first instance, i.e., the approach will not create an alternative, but less-plausible, explanation unless the current best has to be discarded. With regard to the use of the modulus of a derived FCN for plausibility evaluation, r_2 and r_4 are established to outperform the rest. This means that either "the sound volume is *loud* and the distance is *medium*" or "the sound volume is *medium* and the distance is *near*" is the most-plausible

situation to have caused the *quite_high* degree of window being shattered.

From Table II, let us take r_2, for example, $\tilde{z}_{antecedent_{r_2}} = 1 + i1$; it follows from (21) that

$$\text{Re}(\tilde{z}_{antecedent_{r_2}}) = \min(\text{Re}(\tilde{z}_{x_2:loud}), \text{Re}(\tilde{z}_{x_3:medium}))$$

$$\text{Im}(\tilde{z}_{antecedent_{r_2}}) = \min(\text{Im}(\tilde{z}_{x_2:loud}), \text{Im}(\tilde{z}_{x_3:medium})).$$

It is obvious that

$$\text{Re}(\tilde{z}_{antecedent_{r_2}}) = 1 \iff \begin{cases} \text{Re}(\tilde{z}_{x_2:loud}) = 1 \\ \text{Re}(\tilde{z}_{x_3:medium}) = 1 \end{cases}$$

$$\text{Im}(\tilde{z}_{antecedent_{r_2}}) = 1 \iff \begin{cases} \text{Im}(\tilde{z}_{x_2:loud}) = 1 \\ \text{Im}(\tilde{z}_{x_3:medium}) = 1. \end{cases}$$

Hence, the FCNs associated with $x_2 : loud$ and $x_3 : medium$ can be, respectively, written as

$$\tilde{z}_{x_2:loud} = 1 + i1, \tilde{z}_{x_3:medium} = 1 + i1.$$

Similarly, the FCNs attached to r_4 can be derived. After that, the quantity spaces of x_1, x_2, and x_3 have been revised, and these variables are added to the *CHANGED* array in Algorithm 1, thereby continuing the process of scenario-space generation. The newly created instances of plausible nodes are recursively used in the same manner as if each of them was one piece of original evidence. As a result, MF$_3$ and MF$_4$ are then instantiated in sequence, x_4–x_9 are then added, and the FCNs associated with their states can be obtained in the same way as above. This phase of scenario generation (with given e_1) leads to what is shown in Fig. 7. The remaining rule instances can be formed into two global solutions, as presented in Table III. These two solutions are equally suitable as they have the same FCNs attached.

At the termination of the above iteration, MF$_4$, which was previously retrieved by the other two pieces of initial evidence, i.e., e_2 and e_3, is then activated to continue the process of scenario generation until the $CHANGED$ array is empty.In this case, no backward propagation is needed as these two pieces of information only match the antecedent of MF$_4$. The overall matching degree between them has been illustrated (see Section V-C); the FCNs attached to the corresponding scenario nodes are then required to be derived. As an example, the results of matching-rule instance r_1 of MF$_4$ with e_2 and e_3 happen to be H for the imaginary part of each FCN associated with its instantiated variables (i.e., x_5–x_8). Thus, the FCNs for these nodes are $VH + iH$,
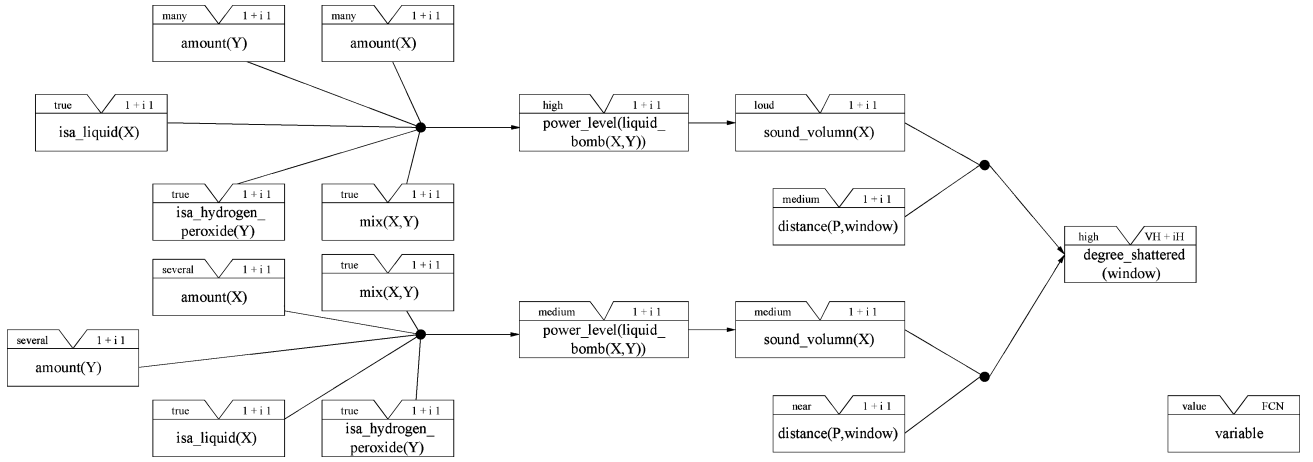
Fig. 7.   Emerging scenario space after initiated by e_1.

TABLE III
TWO PLAUSIBLE SOLUTIONS TO THE CURRENT PROBLEM

Vars	S_1	S_2
x_1	high $\langle 1 + 1i \rangle$	high $\langle 1 + 1i \rangle$
x_2	loud $\langle 1 + 1i \rangle$	medium $\langle 1 + 1i \rangle$
x_3	medium $\langle 1 + 1i \rangle$	near $\langle 1 + 1i \rangle$
x_4	high $\langle 1 + 1i \rangle$	medium $\langle 1 + 1i \rangle$
x_5	many $\langle 1 + 1i \rangle$	several $\langle 1 + 1i \rangle$
x_6	many $\langle 1 + 1i \rangle$	several $\langle 1 + 1i \rangle$
x_7	true $\langle 1 + 1i \rangle$	true $\langle 1 + 1i \rangle$
x_8	true $\langle 1 + 1i \rangle$	true $\langle 1 + 1i \rangle$
x_9	true $\langle 1 + 1i \rangle$	true $\langle 1 + 1i \rangle$

TABLE IV
RESULTS OF FORWARD PROPAGATION

Derived by	Consequent	$Re(\tilde{z}_c)$	$Im(\tilde{z}_c)$
r_1	x_4: high	H	H
r_2	x_4: high	M	M
r_3	x_4: medium	M	M
r_4	x_4: medium	M	L
r_5	x_4: medium	H	M
r_6	x_4: low	M	L
r_7	x_4: low	L	VL

$H + iH$, $VH + iH$, and $H + iH$, respectively. Note that the real part of each of these FCNs is inherited from the certainty degree attached to the given evidence. For instance, for node "x_5 is *many*," its certainty degree is the one assigned to evidence e_2 ("amount of cola is *a_lot*" with a certainty degree of *very_high*). Hence, the FCN associated with this node is $VH + iH$.

The resulting FCNs are those associated with instantiated variables within the individual rule instances of MF$_4$. They are combined using (18) to obtain the overall FCN of the antecedent part of a rule instance. For r_1, the resulting overall FCN is $H + iH (= \min(VH, H, VH, H) + i \min(H, H, H, H))$. From this, applying the forward propagation method, i.e., (20), results in the FCNs associated with the instantiated consequent of each rule instance. These results are listed in Table IV. Results of the above procedure are then passed through the *compatibility filter*. For this simple example,
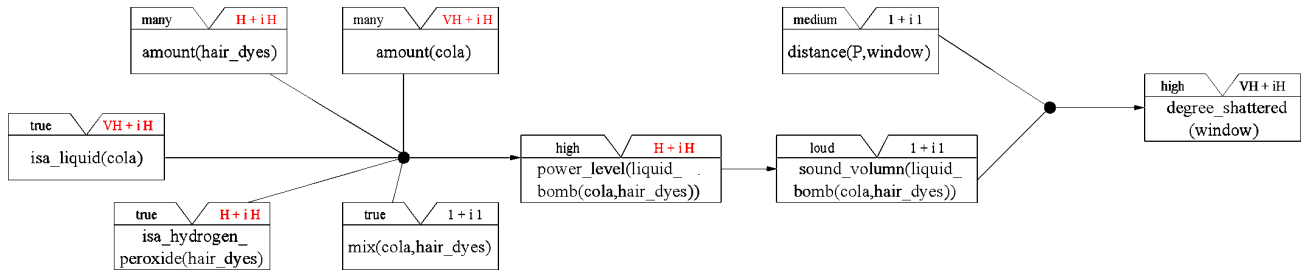
TABLE V
OUTCOME OF PAIRWISE FILTERING

Derived by	Consequent	$Re(\tilde{z}_c)$	$Im(\tilde{z}_c)$
r_1	x_4: high	H	H
r_5	x_4: medium	H	M

no *nogood* MFs are provided in the KB. Thus, no instantiated consequents are removed. After this, the results are processed by the *pairwise filter*. For example, the reasonal consequence of r_1 already exists in the emerging scenario space. The FCN associated with this description must be updated with its existing counterpart using (22). The resulting FCN is (again) $H + iH$. The outcome of pairwise filtering is listed in Table V.

Given Table V, it is straightforward to calculate the moduli of all the FCNs attached to the different consequents of MF$_4$. Clearly, the modulus of FCN of r_1 is of the higher value, which implies that the reasonal consequence of "the power level of the liquid bomb combined by cola and hair_dyes being *high*" requires higher focus of attention. This leads to the focused scenario space, as shown in Fig. 8. For this example, no further instantiations and propagations of variable states are possible. Hence, Fig. 8 is the final (prioritized) outcome of the entire FCM process.

E. Analysis of Generated Scenario Space

It is obvious that many explosive ingredients and liquids can be combined to create homemade liquid bombs. However, there are a lot of explosive chemicals that can be concocted from some very common items, which are otherwise innocent. Under many circumstances, initially collected information or data may seem to be irrelevant, which makes it very difficult for intelligence analysts to detect a plausible threat. In the above example, the initial pieces of evidence may not seen to be very relevant to the explosion case under investigation from the outset. The proposed fuzzy compositional modeler helps in building a scenario space that can offer the best-possible explanation for the given crime scene. The scenario describes the hypothesis that if the suspect mixes the cola and hair dye in an appropriate proportion,

Fig. 8. Emerging (and final) scenario space after propagated by e_2 and e_3.

TABLE VI
CONVENTIONAL CM VERSUS FCM

Task	Approach	
	CM	FCM
Knowledge representation	crisp MFs and prior numerical probability distributions	fuzzy MFs and linguistic uncertainty terms
MF retrieval	boolean retrieval	flexible retrieval
Model composition	iterative abduction and deduction operations in conjunction with Bayesian Networks	Waltz algorithm in conjunction with constraint filtering
Model evaluation	entropy	modulus of FCNs

a liquid bomb can be produced to cause the *quite_high* degree of window shattering. Note that if the generated scenario space contains more than one hypothesis, the plausibility can be globally evaluated with respect to their attached FCNs. The one with the largest modulus of the derived FCN can be first taken into account by the investigators.

Further, the generated scenario space can also be applied to guide further evidence collection. For example, Fig. 8 suggests that the distance between the liquid bomb and the window is *medium*, which indicates that the individuals lying either *near* or *far* from the window are more likely to be victims rather than suspects. In other words, it is more efficient and effective to identify the suspect by searching the dead bodies lying at a *medium* distance away from the window.

F. Comparison With Existing Compositional Modeling Work

This section presents results of comparing the use of different CM techniques for crime investigation. The methods that are employed by different approaches [27] to implement CM are detailed in Table VI. The performance and capability of different approaches within each CM component, as shown in Fig. 1, are discussed as follows.

1) Knowledge Representation: Each assumption or fact in the nonfuzzy-CM approach requires a prior numerical-probability distribution, which must either be predefined in the KB or be specified by the user. However, the degree of precision of available knowledge and data can vary greatly, subject to different perception, judgement, and individuality of people. Thus, it is often very difficult, if not impossible, to obtain such complete probability distributions. Additionally, the num-

ber of required probability measures increases exponentially with respect to the number of involved variables. Considering the above crime scene, for example, in MF_4, a complete probability distribution requires 600 (i.e., $5 \times 5 \times 2 \times 2 \times 2$) numerical-probability values. The embedded knowledge in the existing KB is insufficient to build such probability distributions. For FCM, every possible outcome does not need to be specified, only the most-significant and interesting components are considered and represented in qualitative terms (which are much easier to obtain).

2) Model Fragment Retrieval: In nonfuzzy-CM approach, boolean retrieval is employed to implement the corresponding component. This restrictively requires that both the semantics of a variable and its associated values be matched precisely; otherwise, no retrieval is possible. In the above example, given the collected evidence E and the KB, none of the MFs will be activated for further composition due to no exact match can be made (e.g., no rule instance in MF_2 fully matches the observation of window shattering with the "quite_high" degree). As a result, the crisp CM approach is incapable of detecting and generating any plausible scenarios, while the proposed fuzzy compositional modeler can make useful suggestions.

3) Model Composition: Two conventional inference techniques, i.e., abduction and deduction, are iteratively applied in [27] to develop the scenario space. The number of iterations plays a crucial role in the process of model composition, while it is difficult to automatically specify this number. If the number is too small, it is hard to ensure completeness of the generated scenario space. If, however, the number is too large, it may incur considerable computational overheads. For FCM, due to the employment of the Waltz algorithm (which is proven

computationally very efficient), the generation of a complete model space is ensured during the composition process itself. In fact, the composition process stops automatically when no further changes on the quantity spaces of involved variables are produced.

Existing work on CM for crime investigation involves two separate procedures: A structural scenario space is generated first, and then, a corresponding probabilistic scenario space is computed. A probabilistic scenario space is actually a Bayesian network that describes how likely a certain combination of situations affects other situations. For each node that does not correspond to an assumption or fact, a conditional-probability distribution table based on the probability distribution associated with each MF is calculated. One significant drawback of this technique is its incapability to perform backward propagation of uncertainty. In the above example, given e_1, suppose that MF_2 ($x_2 \wedge x_3 \rightarrow x_1$) is retrieved and a prior probability distribution of x_1 is provided. The probability distributions of x_2 and x_3 cannot be derived due to the lack of inverse calculus of conditional probability. As such, the existing approach only works when the collected evidence appears in the antecedent part of the active MFs. For FCM, the inexact information can be propagated both forward and backward effectively by employing Algorithms 3 and 4. This substantially extends the scope to generate possible scenario descriptions, which is desirable for the purpose of crime investigation.

4) Model Evaluation: The crisp CM work generates all the possible scenarios at the same time, without the ability to prioritize generated scenarios. To differentiate between the likelihood of the scenarios within the constructed probabilistic space, further evidence collection is necessary. For this, entropy is employed as a measure to compute evidence-collection strategies (which supports the selection of an investigation action by minimizing the expected value of the entropy over a set of hypotheses H given evidence E). In contrast, FCM only creates the most-likely scenarios in the first instance. Less-likely ones are generated only when those that are more likely have been refuted. The modulus of the attached FCN is interpreted as a measure to evaluate generated scenarios directly. Hence, it can be concluded that in terms of implementational simplicity, FCM outperforms the crisp CM approach without the need to introduce any additional measures.

In summary, compared with previous work on CM, the performance of FCM is greatly improved with regard to a number of important properties that an automated modeler should possess, including robustness, completeness, simplicity, and interpretability. Of course, FCM requires specification of fuzzy quantities and domain ontology. Its success relies on the quality (such as completeness and coherence) of the predefined KB. Nevertheless, all CM methods are themselves knowledge-based—they all require prior knowledge about the problem domain one way or another. In particular, specification of numerical-probability distributions associated with MFs, as demanded by the existing CM work in crime investigation, is much more difficult to obtain, as compared with qualitative description of linguistic terms about domain variables.

VI. CONCLUSION

This paper has developed an innovative framework of the fuzzy compositional modeler that is capable of automatically generating plausible-scenario descriptions, given inexact knowledge and data. In particular, a novel notion of FCNs has been established and integrated into the existing CM framework, for the first time, to handle 2-D uncertainty explicitly, as well as to constrain the generation of scenarios. The utility and usefulness of the proposed framework are illustrated by means of an application to the construction of plausible descriptions from given evidence in the crime-investigation domain.

While the applicability of the notion of FCN is demonstrated to perform the task of CM, the notion of FCNs and their operations are general and can be readily adapted to suit many different problems, such as performance evaluation [26]. Additionally, from the CM perspective, the current research improves this field both in breadth and in depth. It not only provides a more concise and flexible knowledge-representation formalism for generic MFs and multimodal inexact information but also complements the conventional CM work with regard to the robustness, completeness, simplicity, and interpretability.

Although the proposed approach is promising, much may be done through further research. One such aspect is to extend the automated modeling process in a more dynamic manner, i.e., to generate new scenarios with respect to the changing modeling environment efficiently and effectively, with minimal disruption to the generated scenarios. In addition, a more general constraint-satisfaction mechanism, which is suitable for constraining variables that are modified by FCNs, would help to improve the generality of this work (as, at present, there is a need to prespecify the fuzzy quantities used in FCN-based CM). In particular, this may avoid the need to prefix just one common quantity space from which the real and imaginary parts of all FCNs may take values. Third, weights on the edges in the taxonomical tree are currently numerical. A natural extension is, therefore, to consider the employment of fuzzy numbers to replace the seemingly precise numerical weights. This will, of course, require revision to the semantic similarity calculus. This may lead to ground-breaking techniques for linguistic similarity measures, thereby making a contribution to computing with words [22].

APPENDIX A

KEY MODEL FRAGMENTS IN SAMPLE KNOWLEDGE BASE

Let us define *fuzzyvariable*{

Name: amount   Universe of discourse: [0, 1]

Unit: none   Cardinality of partition: 5

Quantity space:

$$fs_1 = \left[0, 0, \frac{1}{n-1}\right]$$
$$\cdots$$
$$fs_i = \left[\frac{i-2}{n-1}, \frac{i-1}{n-1}, \frac{i}{n-1}\right]$$

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

Names of fuzzy sets: {none, few, several, many, a_lot}
Unifiability: amount(X)}.

Let us define *fuzzyvariable*{
Name: power_level   Universe of discourse: [0, 1]
Unit: none   Cardinality of partition: 3

Quantity space:

$$fs_1 = \left[0, 0, \frac{1}{n-1}\right]$$
$$\cdots$$
$$fs_i = \left[\frac{i-2}{n-1}, \frac{i-1}{n-1}, \frac{i}{n-1}\right]$$
$$\cdots$$
$$fs_n = \left[\frac{n-2}{n-1}, 1, 1\right]$$

Names of fuzzy sets: {low, medium, high}
Unifiability: power_level(X)}.

Let us define *fuzzyvariable*{
Name: sound_volume   Universe of discourse: [0, 1]
Unit: decibels   Cardinality of partition: 3

Quantity space:

$$fs_1 = \left[0, 0, \frac{1}{n-1}\right]$$
$$\cdots$$
$$fs_i = \left[\frac{i-2}{n-1}, \frac{i-1}{n-1}, \frac{i}{n-1}\right]$$
$$\cdots$$
$$fs_n = \left[\frac{n-2}{n-1}, 1, 1\right]$$

Names of fuzzy sets: {low, medium, loud}
Unifiability: sound_volumn(X)}.

Let us define *fuzzyvariable*{
Name: distance   Universe of discourse: [0, 1]
Unit: kilometer   Cardinality of partition: 3

Quantity space:

$$fs_1 = \left[0, 0, \frac{1}{n-1}\right]$$
$$\cdots$$
$$fs_i = \left[\frac{i-2}{n-1}, \frac{i-1}{n-1}, \frac{i}{n-1}\right]$$
$$\cdots$$
$$fs_n = \left[\frac{n-2}{n-1}, 1, 1\right]$$

Names of fuzzy sets: {near, medium, far}
Unifiability: distance(P,X)}.

Let us define *fuzzyvariable*{
Name: degree_of_shattered   Universe of discourse: [0,1]
Unit: none   Cardinality of partition: 5

Quantity space:

$$fs_1 = \left[0, 0, \frac{1}{n-1}\right]$$
$$\cdots$$
$$fs_i = \left[\frac{i-2}{n-1}, \frac{i-1}{n-1}, \frac{i}{n-1}\right]$$
$$\cdots$$
$$fs_n = \left[\frac{n-2}{n-1}, 1, 1\right]$$

Names of fuzzy sets: {very_low, low, medium, quite_high, high}
Unifiability: degree_of_shattered(X)}

IF {person(P), substance(X), location(L)}
Assuming {possess(P, X)}
THEN {seen_with(P, X, L)}
Distribution seen_with(P, X, L){
r_1: true, true, true, true→true:H,
r_2: true, true, true, true→false:VL} (MF$_1$)

IF {sound_volumn(X), distance(P,Y)}
THEN {degree_of_shattered(Y)}
Distribution degree_of_shattered(Y){
r_1: loud, near → high:VH
r_2: loud, medium → high:H
r_3: loud, far → high:L
r_4: medium, near → high:H
r_5: medium, medium → high:M
r_6: medium, far → high:L
r_7: low, near → high:M
r_8: low, medium → high:L
r_9: low, far → high:VL} (MF$_2$).

IF {power_level(X)}
THEN {sound_volumn(X)}
Distribution sound_volumn(X){

r_1: low → low:VH,	r_2: low → medium:M
r_3: low → loud:VL	r_4: medium → low:L
r_5: medium → medium:VH	r_6: medium → loud:L
r_7: high → low:VL	r_8: high → medium:M
r_9: high → loud:VH} (MF$_3$).	

IF {amount(X), amount(Y), isa_liquid(X)
isa_hydrogen_peroxide(Y)}
Assuming {mix(X, Y)}
THEN {power_level(liquid_bomb(X, Y))}.
Distribution power_level(liquid_bomb(X, Y)){
r_1: many, many, true, true, true, true → high:VH
r_2: many, several, true, true, true, true → high:M
r_3: many, few, true, true, true, true → medium:M
r_4: several, many, true, true, true, true → medium:M

r_5: several, several, true, true, true, true → medium:VH
r_6: several, few, true, true, true, true → low:H
r_7: few, few, true, true, true, true → low:L} (MF$_4$).

Collected Evidence

e_1: degree_of_shattering(window) = quite_high with certainty degree: VH

e_2: amount(cola) = a_lot with certainty degree: VH

e_3: amount(hair_dyes) = several with certainty degree: H.

REFERENCES

[1] T. Boongoen, Q. Shen, and C. Price, "Disclosing false identity through hybrid link analysis," *Artif. Intell. Law*, 2010. DOI: 10.1007/s10506-010-9085-9.

[2] J. J. Buckley, "Fuzzy complex numbers," *Fuzzy Sets Syst.*, vol. 33, no. 3, pp. 333–345, 1989.

[3] J. J. Buckley and Y. Qu, "Fuzzy complex analysis I: Differentiation," *Fuzzy Sets Syst.*, vol. 41, no. 3, pp. 269–284, 1991.

[4] J. J. Buckley, "Fuzzy complex analysis II: Integration," *Fuzzy Sets Syst.*, vol. 49, no. 2, pp. 171–179, 1992.

[5] G. de Cooman, "A behavioural model for vague probability assessments," *Fuzzy Sets Syst.*, vol. 154, no. 3, pp. 305–358, 2005.

[6] S. Dick, "Toward complex fuzzy logic," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 405–414, Jun. 2005.

[7] B. Falkenhainer and K. Forbus, "Compositional modelling: Finding the right model for the job," *Artif. Intell.*, vol. 51, pp. 95–143, 1991.

[8] X. Fu, T. Boongoen, and Q. Shen, "Evidence directed generation of plausible crime scenarios with identity resolution," *Appl. Artif. Intell.*, vol. 24, no. 4, pp. 253–276, 2010.

[9] X. Fu and Q. Shen, "Fuzzy model fragment retrieval," in *Proc. 17th Int. Conf. Fuzzy Syst.*, 2008, pp. 1381–1388.

[10] X. Fu and Q. Shen, "A novel framework of fuzzy complex numbers and its application to compositional modelling," in *Proc. 18th Int. Conf. Fuzzy Syst.*, 2009, pp. 536–541.

[11] X. Fu, Q. Shen, and R. Zhao, "Towards fuzzy compositional modelling," in *Proc. 16th Int. Conf. Fuzzy Syst.*, 2007, pp. 1233–1238.

[12] V. J. Geberth, *Practical Homicide Investigation*, 4th ed. Boca Raton, FL: CRC, 2006.

[13] M. Ha, W. Pedrycz, and L. Zheng, "The theoretical fundamentals of learning theory based on fuzzy complex random samples," *Fuzzy Sets Syst.*, vol. 160, no. 17, pp. 2429–2441, 2009.

[14] J. Halliwell and Q. Shen, "Linguistic probabilities: Theory and application," *Soft Comput.*, vol. 13, no. 2, pp. 169–183, 2009.

[15] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[16] J. Keppens and Q. Shen, "On compositional modelling," *Knowl. Eng. Rev.*, vol. 16, no. 2, pp. 157–200, 2001.

[17] J. Keppens and Q. Shen, "Compositional model repositories via dynamic constraint satisfaction with order-of-magnitude preferences," *J. Aritif. Intell. Res.*, vol. 21, pp. 499–550, 2004.

[18] M. Koyuncu and A. Yazici, "A fuzzy knowledge-based system for intelligent retrieval," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 317–330, Jun. 2005.

[19] K. R. Lang, S. Ba, and A. B. Whinston, "Compositional enterprise modeling and decision support," *Handbook Decis. Support Syst.*, vol. 2, pp. 1443–1468, 2008.

[20] R. Leitch, M. Chantler, Q. Shen, and G. Coghill, "A preliminary specification methodology for model-based diagnosis," *Ann. Math. Artif. Intell.*, vol. 11, pp. 11–32, 1994.

[21] K. S. Leung and W. Lam, "A fuzzy expert system shell using both exact and inexact reasoning," *J. Autom. Reason.*, vol. 5, no. 2, pp. 207–233, 1989.

[22] J. M. Mendel, "Computing with words and its relationships with fuzzistics," *Inf. Sci.*, vol. 177, no. 4, pp. 988–1006, 2007.

[23] J. Qiu, C. Wu, and F. Li, "On the restudy of fuzzy complex analysis: Part I. The sequence and series of fuzzy complex numbers and their convergences," *Fuzzy Sets Syst.*, vol. 115, no. 3, pp. 445–450, 2000.

[24] J. Qiu, C. Wu, and F. Li, "On the restudy of fuzzy complex analysis: Part II. The continuity and differentiation of fuzzy complex functions," *Fuzzy Sets Syst.*, vol. 120, no. 3, pp. 517–521, 2001.

[25] D. Ramot, R. Milo, M. Friedman, and A. Kandel, "Complex fuzzy sets," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 171–186, Apr. 2002.

[26] K. A. Rasmani and Q. Shen, "Data-driven fuzzy rule generation and its application for student academic performance evaluation," *Appl. Intell.*, vol. 25, no. 3, pp. 305–319, 2006.

[27] Q. Shen, J. Keppens, C. Aitek, B. Schafer, and M. Lee, "A scenario driven decision support system for serious crime investigation," *Law, Probability Risk*, vol. 5, no. 2, pp. 87–117, 2006.

[28] Q. Shen and R. Leitch, "Fuzzy qualitative simulation," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 4, pp. 1038–1061, Apr. 1993.

[29] P. Walley, "Measures of uncertainty in expert systems," *Artif. Intell.*, vol. 83, pp. 1–58, 1996.

[30] D. Waltz, "Understanding line drawings of scenes with shadows," in *The Psychology of Computer Vision*, P. H. Winston, Ed. Cambridge, MA, McGraw-Hill, 1975, pp.19–91.

[31] G. Wang and X. Li, "Generalized Lebesgue integrals of fuzzy complex valued functions," *Fuzzy Sets Syst.*, vol. 127, no. 3, pp. 363–370, 2002.

[32] C. Wu and J. Qiu, "Some remarks for fuzzy complex analysis," *Fuzzy Sets Syst.*, vol. 106, no. 2, pp. 231–238, 1999.

[33] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 100, no. Supp., pp. 9–34, 1999.

[34] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, 1975.

[35] G. Zhang, "Fuzzy limit theory of fuzzy complex numbers," *Fuzzy Sets Syst.*, vol. 46, no. 2, pp. 227–235, 1992.

Xin Fu (S'07) received the B.Sc. degree (first-class honors) in computer science from the University of Ulster, Coleraine, U.K., in 2005 and the M.Phil. degree in informatics from the University of Manchester, Manchester, U.K., in 2007. She is currently working toward the Ph.D. degree in computer science at Aberystwyth University, Aberystwyth, U.K.

Her research interests include fuzzy and qualitative modeling and intelligent decision support for crime detection and prevention.

Qiang Shen received the B.Sc. and M.Sc. degrees in communications and electronic engineering from the National University of Defence Technology, China, and the Ph.D. degree in knowledge-based systems from Heriot-Watt University, Edinburgh, U.K.

He holds the established Chair in computer science and is currently the Head of the Department of Computer Science, Aberystwyth University, Aberystwyth, U.K. He is also an Honorary Fellow with the University of Edinburgh, Edinburgh, U.K. He has authored two research monographs. He has authored or coauthored more than 250 peer-reviewed papers. He has chaired and given keynote lectures at many international conferences. His current research interests include computational intelligence, fuzzy and qualitative modeling, reasoning under uncertainty, pattern recognition, data mining, and real-world applications of such techniques for decision support (e.g., crime detection, consumer profiling, systems monitoring, and medical diagnosis).

Prof. Shen is currently an Associate Editor of two premier IEEE TRANSACTIONS (i.e., the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, and the IEEE TRANSACTIONS ON FUZZY SYSTEMS). He is also an Editorial Board Member of several other leading international journals. He received an Outstanding Transactions Paper Award from the IEEE for a peer-reviewed paper.