Representation and Reasoning Under Uncertainty in Deception Detection: A Neuro-Fuzzy Approach

Lina Zhou, *Member, IEEE*, and Azene Zenebe

***Abstract*—An analysis of the process and human cognitive model of deception detection (DD) shows that DD is infused with uncertainty, especially in high-stake situations. There is a recent trend toward automating DD in computer-mediated communication. However, extant approaches to automatic DD overlook the importance of representation and reasoning under uncertainty in DD. They represent uncertain cues as crisp values and can only infer whether deception occurs, but not to what extent deception occurs. Based on uncertainty theories and the analyses of uncertainty in DD, we propose a model to represent cues and to reason for DD under uncertainty, and address the uncertainty due to imprecision and vagueness in DD using fuzzy sets and fuzzy logic. Neuro-fuzzy models were developed to discover knowledge for DD. The evaluation results on five data sets showed that the neuro-fuzzy method not only was a good alternative to traditional machine-learning techniques but also offered superior interpretability and reliability. Moreover, the gains of neuro-fuzzy systems over traditional systems became larger as the level of uncertainty associated with DD increased. The findings of this paper have theoretical, methodological, and practical implications to DD and fuzzy systems research.**

***Index Terms*—Deception detection, fuzzy sets, fuzzy systems, uncertainty.**

I. Introduction

DECEPTION occurs when someone knowingly transmits information to create false conclusions [1]. Deception is when an unidentified source sends an e-mail warning of potential security breaches to an online banking system asking you to log into your bank account via a provided link. Deception detection (DD) entails high stakes to both individuals and organizations. For example, successful DD may prevent an investor from suffering serious financial loss or keep a criminal from getting loose. However, DD is a challenging task due to the convergence of a variety of factors such as communication modalities and an individual's communication skills. As a result, humans are not good detectors. Recent advances in automatic DD have produced promising results and have enabled DD to be performed faster than before (e.g., [2]). Nonetheless, the result of automatic DD is still far from satisfying humans' practical needs. There is a pressing need to improve the state of DD by advancing the knowledge and techniques for DD and inhancing the decision support for human DD.

Manuscript received March 23, 2005; revised August 14, 2005 and January 24, 2006. This work was supported in part by the National Science Foundation (NSF) under Grant 0328391 and by UMBC under a Special Research Initiative Support/Research Assistant Grant.

L. Zhou is with the Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: zhoul@umbc.edu).

A. Zenebe is with the Department of Management Information Systems, Bowie State University, Bowie, MD 20715 USA (e-mail: azenebe@bowiestate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2006.889914

In recent years, several studies on automatic DD [2], [3] have emerged. The studies apply traditional machine-learning techniques by considering both cues to deception and DD outcomes as variables with crisp values. A number of cues to deception have been found to have strong effects across multiple studies [4]–[6], some of which have been tested with real-world data [7], [8]. For example, quantity (of messages) is an effective cue to deception because deceivers become verbose during persuasion in computer-mediated communication [4], [6]. In a traditional machine-learning model for DD, quantity is measured by the number of words in utterances or messages and the final predictions are either deceptive or truthful. The primary goal of the above approach is to achieve high precision, certainty, and rigor. It has been revealed from previous deception research, however, that cues to deception are subject to a variety of moderating factors (e.g., objectively versus subjectively measured cues and duration of presentation) [4], and adaptation between communication partners can occur during deceptive interactions [9]–[11]. These findings highlight the uncertain nature of DD.

The uncertainty refers to epistemic uncertainty induced from imperfection of information including imprecision, vagueness, nonspecificity, and conflict [12]. For example, uncertainty can be induced by imprecise observation and characterization of cues to deception as well as linguistic ambiguity and vagueness in messages. Uncertainty is present in many subjective cues such as pleasantness and activation, and some linguistic cues such as lexical diversity. Such cues post challenges to representing and discerning their values ranging from representing and discerning the values obtained for such cues ranging from very high to very low and inferring their contributions to the final DD outcome. Deception rarely takes the form of blatant lies. In order to avoid being detected, a deceiver may occasionally take low-key status and speak the truth [11]. Zadeh [13] argues that everything including truth is a matter of degree. When it comes to making a detection decision, people tend to struggle between two extreme choices: (definitely) deceptive and (definitely) truthful. Thus, DD is carried out under uncertainty. By examining uncertainty, we would be able to tell not only whether deception occurs but also to what extent deception occurs.

Given serious ethical and even legal consequences involved with DD, automatic DD is most likely to be used as decision support for humans rather than as the final decision maker, especially in high-stake situations. Decision support is typically a knowledge-based system that is expected to provide explanations to help overcome comprehension difficulties caused by various types of perceived anomalies in the output [14]. Based on the cognitive model involved in decision making, it would be more comfortable for people to choose between alternative outcomes with different degrees of possibilities than accept a statement being definitely deceptive or truthful. Such humans' ability to tolerate uncertainty has yet to be studied and exploited in automatic DD.

Uncertainty in DD refers to the inability to discern whether someone is telling the truth or being deceptive. The importance of uncertainty to DD lies in the fact that we do not have complete information about an object or event, or about the knowledge of deception cues that would help us decide whether a statement is true or false. We make the first effort to examine uncertainty in automatic DD in this paper. In particular, we aim to answer the following research question: how do we represent and reason under uncertainty in automatic DD?

We conducted a systematic analysis of uncertainty associated with DD. Fuzzy sets exploit uncertainty in an attempt to make system complexity manageable. Fuzzy logic is not only able to deal with incomplete, noisy, or imprecise data but also helpful in developing uncertainty models of the data that provide smarter and smoother performance than do traditional systems [15, p. 256]. Based on fuzzy set theory, we developed a computational model to represent and infer uncertainty induced by imprecision and vagueness in cues to DD, their relationships, and the resulting outcomes. Neuro-fuzzy models, fuzzy systems enhanced by neural network learning, were developed to determine the fuzzy sets of cues to deception and to discover knowledge for DD. We compared the performance of neuro-fuzzy models against two traditional machine-learning models—decision trees and neural networks. Data sets were collected from five types of communication settings. The results showed that the proposed models not only were capable of representing uncertainty in DD but also produced comparable and even better accuracy. In addition, the neuro-fuzzy approach improved the interpretability, stability, and reliability of DD models.

The contributions of this paper to fuzzy systems research are the following.

a) We designed fuzzy systems in a new application domain. While fuzzy systems are well studied for control systems in engineering, this is the first time that fuzzy systems are applied to DD in interpersonal communication.
b) We laid the theoretical foundation for modeling and reasoning uncertainty due to imprecision and vagueness in DD by using fuzzy theories and systems.
c) We advanced the development of knowledge-based fuzzy systems by automatically learning rules from data, which alleviates the knowledge bottleneck problem.
d) We provided empirical evidence showing that fuzzy rules are easy to interpret and neuro-fuzzy models are stable and less sensitive to imprecision across different communication modalities and tasks for DD.

The contributions of this paper to deception research and practice are the following.

a) We created an uncertainty model for DD. The model can be used to not only guide the development of decision support systems for human DD but also improve our understanding of the process of human DD.
b) We identified uncertainty types associated with DD (e.g., uncertainty due to imprecision and vagueness), which lay the groundwork for addressing uncertainty in DD.
c) We developed approaches to representing uncertainty in cues to deception and to reasoning uncertainty in the outcome of DD.
d) We built neuro-fuzzy models for automatic DD. The discovered knowledge can also be used to train humans in manual DD as well as serve as decision aids in support of human DD. This is especially promising for DD in computer-mediated communication, where there is a lack of prior knowledge and domain experts, and text is the major source of information.

The rest of this paper is organized as follows. In Section II, we provide the theoretical foundation for uncertainty in DD. We also propose a model to represent uncertainty in DD and a computation model to support automatic DD. In Section III, we introduce a methodology to represent and reason imprecision and vagueness in DD. We report experiment results and discuss findings and their implications to DD research and practice in Section IV. Finally, in Section V, we highlight the contributions of this paper and suggest research questions that warrant further investigation.

II. Uncertinty in Deception Detection (DD)

Uncertainty presents itself in knowledge-based systems [16]. For example, uncertainty has been studied in the context of multiagent systems [17] for intrusion detection and reputation management [18]. However, uncertainty in DD in interpersonal communication remains unexplored.

A. Deception Detection Under Uncertainty

Deception in everyday life is achieved by messages that combine truthful and deceptive information (e.g., equivocation, exaggeration) or merely omit relevant details from truthful communication [1]. Given incomplete, imprecise, and vague input information, it is difficult to detect deception accurately, and uncertainty is evident in DD. Traditional DD studies have followed the absolute binary-valued paradigm [2], [3], [19], [20]. Their goal is to make a crisp truthful–deceptive judgment, which can only tell whether deception occurs. The classic logic is ternary, which allows us to represent the unknown in addition to surely deceptive and surely truthful. Uncertainty theories go even further by allowing a variable, be it input or output, to be multivalued with differing degrees. For example, the fuzzy approach would be able to tell how much a message is deceptive and how much it is truthful. Grounded to uncertainty and deception theories and findings, we analyze uncertainty associated with DD and propose models to address such uncertainty.

According to interpersonal deception theory [1], deception is an interplay of the deceptive goal and perceived suspicion on the receiver's side. Suspicion falls somewhere in the intermediate ranges along a truth–falsity judgment continuum

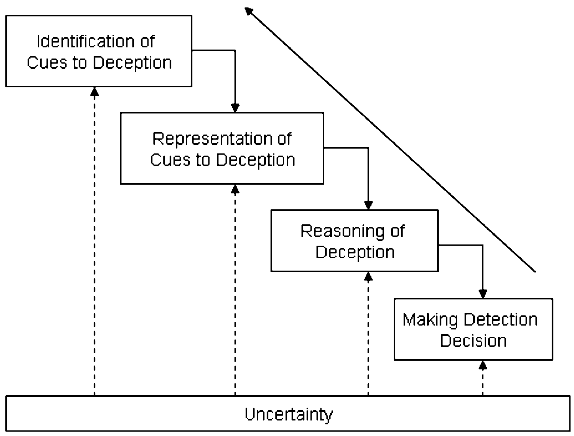

Fig. 1. Uncertainty-infused deception detection (updated from [24]).

[1]. In other words, a suspicious receiver is uncertain about whether the sender is telling the truth or deceiving. This also applies to observers or the third parties who are involved in DD. On the deceiver's side, deception success implicates at least two conflicting goals: achieve deception and avoid being detected. Once the sender decides to deceive, he/she must be concerned about appearing credible, allaying receiver suspicions, minimizing his/her responsibility for deceit, and avoiding unpleasant consequences should deception be detected [1]. Driven by the mixed goals, a deceiver is engaged in both strategic and nonstrategic behaviors during interpersonal communication [1]. The strategic behaviors manifest plans and intentions to achieve deceptive goals, whereas nonstrategic behaviors reflect perceptual, cognitive, and emotional processes accompanying the communicative situation, which are beyond intentional control. Therefore, the theory supports and implies that uncertainty is present and should be accounted for in DD.

DD can be considered as a multiattribute decision-making problem [21], which concerns the choices between two or more alternatives for detection outcomes (e.g., deceptive and truthful), deception detector, state of statement, and costs or payoffs. Each of the alternative outcomes is evaluated with multiple criteria (i.e., cues to deception) and deception tolerance of individuals. One kind of uncertainty in multiattribute decision making arises out of noise or inaccurate criteria evaluation [22]. The vagueness in the focal task of decision making necessarily entails a variety of uncertainty [23]. In terms of state of knowledge in cues to deception and the criteria for making DD decisions, DD is a decision-making problem that is addressed under uncertainty.

B. Uncertainty-Infused Deception Detection Process

The process of DD commonly consists of four phases: cue identification, cue representation, deception reasoning, and making a detection decision. As shown in Fig. 1, DD goes through an iterative process (upward arrow), and uncertainty permeates through the entire course of DD. The omnipresence of uncertainty in DD reinforces its iterative nature. We introduce each phase of DD and its associated uncertainty in the following.

1) Uncertainty in the Identification of Cues to Deception: The task of this phase is to identify cues to deception from a potentially deceptive instance. About 160 cues have been found to indicate deception in a variety of communication settings [4]. Nonetheless, some cues commonly used in face-to-face communication were found to be inaccessible and even show opposite trends in computer-mediated communication [6]. For example, there is a lack of nonverbal cues such as voice pitch and facial expression [17] in computer-mediated communication. Deceivers, who were traditionally considered to be less talkative or productive, were found to be more productive in asynchronous computer-mediated communication [6]. Moreover, it is difficult to judge whether a cue is actually present and how frequently it occurs when it comes to the cue. Therefore, the lack of a complete list of cues to deception and knowledge about the efficacy of each cue leads to uncertainty in DD.

2) Uncertainty in the Representation of Cues to Deception: During this phase, identified cues are formalized and their values measured for further reasoning. For example, a cue can be represented as a symbolic variable or as a variable. Symbolic variables such as pleasantness are vague by nature. Numeric variables are measured by approximation at best. For instance, lexical diversity can be derived by dividing the number of unique words by the total number of words. The count of unique words is contingent upon reducing words to their stem forms by removing inflections (e.g., deceives and deceiving are considered to share the same word token), which in turn introduces imprecision into the measurement of lexical diversity. Another example is pause duration (between messages), which may overlap with the time for composing messages in instant messaging [5]. In addition, uncertainty may be induced by mapping between numerical and symbolic values. Therefore, the approximate or proxy measurement of the cues results in imprecise information for DD.

3) Uncertainty in Deception Reasoning: Inference about deception is conducted based on represented cues during this phase. Imprecise and vague cues to deception induce nonstochastic uncertainty, as defined above, which in turn lead to reasoning under uncertainty. During uncertainty reasoning, techniques are selected to combine cues to deception to derive alternative outcomes along with their possibilities. For example, if cues are represented with fuzzy sets and associated memberships, the reasoning process may arrive at DD outcomes with different possibilities depending on the operators and constraints being applied.

4) Uncertainty in Making Detection Decision: Based on the inference results, a decision is made as to whether a case is deceptive or not. Deception outcomes are subjective, imprecise, and vague. Given the fuzzy nature of deception, it is practically infeasible to reach a decision with 100% certainty. Different individuals and situations have different levels of tolerance for false positive and false negative. For example, given an instance that has a membership of 0.55 in the truthful class and a membership of 0.37 in the deceptive class, respectively, it may be regarded as truthful in one situation and deceptive in another. Thus, the detection decision is uncertain rather than definite.

C. Types of Uncertainty in DD

There are different types of uncertainty based on the underlying causes. According to information theory literature [12],

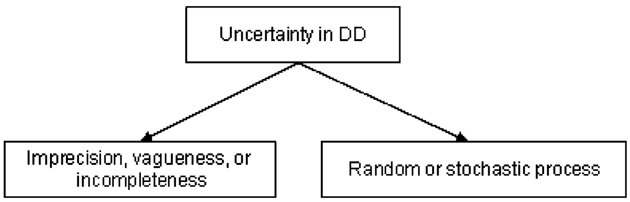

Fig. 2. Sources of uncertainty in DD.

different types of uncertainty are complementary and may coexist in a knowledge-based system. Drawing on the characteristics of DD and the literature on uncertainty theories [25]–[27], [15], we identified two major sources of uncertainty in DD: 1) imprecision, vagueness, or incompleteness and 2) random or stochastic process, as shown in Fig. 2. This uncertainty due to incomplete data and uncertain evidence is referred to as objective uncertainty [15].

1) Imprecision and Vagueness: They result from the difference between the quantity of interest and the proxy that is generally measured. Imprecision and vagueness are induced not only from imprecise observation and categorization of cues to deception and different detection outcomes but also from vagueness in describing cues using linguistic categories such as high and low. In other words, the original source of information and the final outcome of DD are imprecise, vague, and not clear-cut. For instance, given an e-mail message, to what degree is it deceptive or truthful? What does it suggest of deception if the message is high in lexical diversity or contains ten self-references?

2) Incompleteness: Incompleteness is due to inherent properties of the system being studied and our inability to measure them. Given that computer-mediated communication (e.g., e-mails and instant messaging) is leaner than face-to-face communication [28], some cues to deception discovered in face-to-face communication (e.g., vocal tension, fidgeting, leg movements, and posture shifts) are missing in computer-mediated communication. Even in face-to-face communication, an information receiver may not have the capacity to attend to all possible channels of cues at one time. Therefore, incomplete observation of an information receiver or information deficiency creates uncertainty in DD.

3) Random or Stochastic Process: When the information about a future random event is unknown but there is a need to estimate or predict the event, uncertainty results from randomness. Such a type of uncertainty is manifested in information cues randomly displayed in individuals' deception and the presence of random contextual factors in DD. For example, given that the current message is deceptive, it is random to predict the probability of the next message being deceptive.

According to the above discussion, we identified the types of uncertainty associated with individual phases of DD. We also analyzed the causes of uncertainty, as listed in Table I. For example, uncertainty in the identification of cues to deception is caused by the variation in communication context and the missing/incompleteness of important cues to deception.

The analysis of uncertainty issues shed light on a problem with DD. Creating crisp categories by using imprecise measures for vague concepts or incomplete cues in prior DD research ignores vague boundaries that inherently exist between different values of a cue to deception. This highlights the need for a framework that can handle uncertainty in DD to guide future research.

TABLE I
TYPES AND CAUSES OF UNCERTAINTY IN DD PHASES

Phases of DD	Uncertainty types	Causes
Identification	Incompleteness	Variation in communication context and accessibility of cues to deception
Representation	Imprecision and vagueness	Measures are imprecise, approximate, and they approximate vague concepts by nature
Reasoning	Imprecision and vagueness	The nature and representation of cues to deception
Decision making	Imprecision and vagueness Randomness	Outcomes are not clear-cut (i.e., partially deceptive or truthful)

D. A Computational Model for DD Under Uncertainty

Computational models underlie the design and development of knowledge-base systems for automatic DD. It is important for the models to have the ability to represent and reason uncertainty in DD.

A few studies have applied classical machine-learning techniques (e.g., decision trees and neural networks) to DD [2], [3]. Despite promising results, the above techniques showed the following limitations.

a) They are based on the assumption of no overlapping instances, which does not hold for real-world DD applications. For example, in a crisp tree, the cut-point test performs poorly on examples with attribute values close to the cut point [29], [30]. Various types of uncertainty associated with DD are not well supported.
b) They require a large data set compared to fuzzy classifiers [31].
c) Crisp decision trees generate local decisions that may lead to overfitting of the data and high variance due to the fast decrease of local growing sets [29], [32]. Empirical studies show that crisp decision trees produce worse performance than soft decision trees and reduce stability for very small learning sets [29]. The findings pose a problem to DD because deception data sets tend to be small-sized.
d) Despite the strong learning ability of neural networks, they usually involve computational and interpretational complexities [2], [33], [34]. As a result, it is difficult to transform the learned deception models into knowledge that can be directly applied to human DD.

The above limitations could be addressed with a computational model that is able to exploit uncertainty in DD. To this end, we proposed a DD model that extends the discussion of uncertainty in knowledge-based systems [35], [36].

The model is able to discover knowledge for DD under uncertainty. The combination of cues and reasoning in the model should be free of doubtful assumptions such as independence and exclusiveness. The vagueness and imprecise nature of cues to deception and detection outcomes make the use of fuzzy sets and logic theories very promising. The random nature of deception and DD decisions calls for the use of probabilistic measures within a crisp set as well as probabilistic theory to model uncertainty. We can therefore represent uncertainty in DD with two types of measures—fuzzy and probabilistic. Both measures support reasoning under uncertainty and propagation of uncertainty in DD. Since handling multidimensional uncertainty in information systems in a unified way is not well advanced [12], we address uncertainty due to imprecision and vagueness in DD in this paper.

III. Methodology

In this section, we introduce the method for representing and reasoning the uncertainty due to imprecision and vagueness in DD. In addition, we describe experiment design for model evaluation.

A. Representation and Reasoning Uncertainty in DD With Fuzzy Theory

As discussed in Section II, cues to deception are imprecise and vague. Compared with crisp sets, fuzzy sets provide a better framework for cue representation. Specifically, approximate reasoning based on fuzzy theory has received success in control systems engineering and pattern recognition. It is logically valid to use fuzzy sets and fuzzy logic for associated reasoning in making detection decisions because it matches how a human decision maker completes the task.

Fuzzy set theory is basically a theory of classes with unsharp boundaries [37]. It is a computational approach to human reasoning and behavior [13]. Fuzzy sets can include imprecise and vague knowledge based on expert opinions and/or experiences. They provide an approach to representing DD rules that is easy to interpret and use. A general form of fuzzy rules is represented as follows:

$$\text{IF } X_1 \text{ is } \mathrm{A}_{\mathrm{j1}}^{(1,j)} \text{ AND } X_2 \text{ is } \mathrm{A}_{\mathrm{j2}}^{(2,j)} \ldots \text{ AND } X_n \text{ is } \mathrm{A}_{\mathrm{jn}}^{(\mathrm{n},j)} \text{ THEN } X = (X_1 \ldots X_n) \text{ is in } \mathrm{C_k}$$

where X is a vector of the input variables, $\mathrm{A}_{\mathrm{ji}}^{(ij)}$ $(i = 1 \ldots n)$ is a linguistic term of X_i represented as a fuzzy set in rule j, and $\mathrm{C_k}$ represents class k.

For instance, a fuzzy rule for judging a message being deceptive (adapted from a crisp rule in [2]) is represented as follows: If modal verb is large, individual reference is small, and nonspecificity is large, then the message is likely to be deceptive.

Based on the tests on both synthetic and real data sets, fuzzy classifiers have shown performance that is at least comparable to traditional classifiers [31], [38]. Compared with classic machine-learning approaches, a fuzzy (rule-based) method shows the following advantages.

a) Tolerance of vagueness in set membership functions, which may offer robust noise-tolerance models or predications in situations where precise input is unavailable or too expensive [39]. The semantics of a model parameter and encoding of the fuzzy set can often be estimated based on a guess about its probable semantics [40].
b) Less sensitive to precise measurement of membership functions (e.g., cues to deception) and able to model qualitative properties of a phenomenon [41] (e.g., DD).
c) Allows an instance to be classified into more than one target class with varying degrees of membership [13]. For instance, an instance can be classified into truth and deception with degrees of membership of 0.75 and 0.25, respectively. It provides information about uncertainty levels in addition to class labels. By softening the split threshold, noise influence can become less significant in generalizing the model [29].
d) Increases the chance of uncovering hidden knowledge by allowing vagueness [15]. Given the lack of domain experts and prior knowledge on DD, especially in computer-mediated communication, the ability to discover knowledge is very important.

B. Learning DD With Neuro-fuzzy Approach

The neuro-fuzzy approach enhances fuzzy sets with the ability to adapt and learn from experience. It employs a supervised learning algorithm to train fuzzy sets and linguistic rules by modifying the structure and parameters of a neuro-fuzzy model (i.e., inclusion or deletion of neurons or adaptation of the weights). Furthermore, the approach does not hold other assumptions about feature distributions (e.g., normal and random) except for an implicit independence assumption.

Among a variety of neuro-fuzzy tools, we selected NEFCLASS [42] to build DD models. The architecture of the system is shown in Fig. 3. The major features of NEFCLASS include [42] the following.

a) It consists of an input layer $(X_i, i = 1 \ldots M)$, a hidden layer $(R_j, j = 1 \ldots L)$, and an output layer $(C_k, k = 1 \ldots N)$, representing input variables, fuzzy rules, and output variables, respectively. An input node X_i and a rule node R_j are connected by weights $\mathrm{A}_s^{(i,j)}$, where index s represents the selected fuzzy set of the partition [42]. It follows feedforward neural network architecture, while the output is computed by a maximum operation rather than a weighted sum.
b) It incorporates semantics of the underlying fuzzy model into the training process, which preserves the linguistic interpretability of the model. Thus, the black-box behavior of artificial neural networks is avoided. Additionally, it allows shared weights on some of the connections to guarantee that, for each linguistic value of an input variable, there is only one representation as a fuzzy set upon the completion of training. In other words, each fuzzy set that is connected to more than one rule node has one shared weight. For example, two connections in Fig. 3, $\mathrm{A}_3^{(1,j)}$ and $\mathrm{A}_3^{(1,L)}$, both share the third linguistic value of variable X_1.

Neuro-fuzzy learning mainly consists of two steps: creating the structure of a classifier and learning the optimal parameters. The former begins by specifying the number and types of fuzzy sets for each input variable. The latter focuses on training the fuzzy sets. The learning is an iterative process in which the fuzzy sets are adapted using supervised learning from input instances until a number of admissible classifications are reached or the

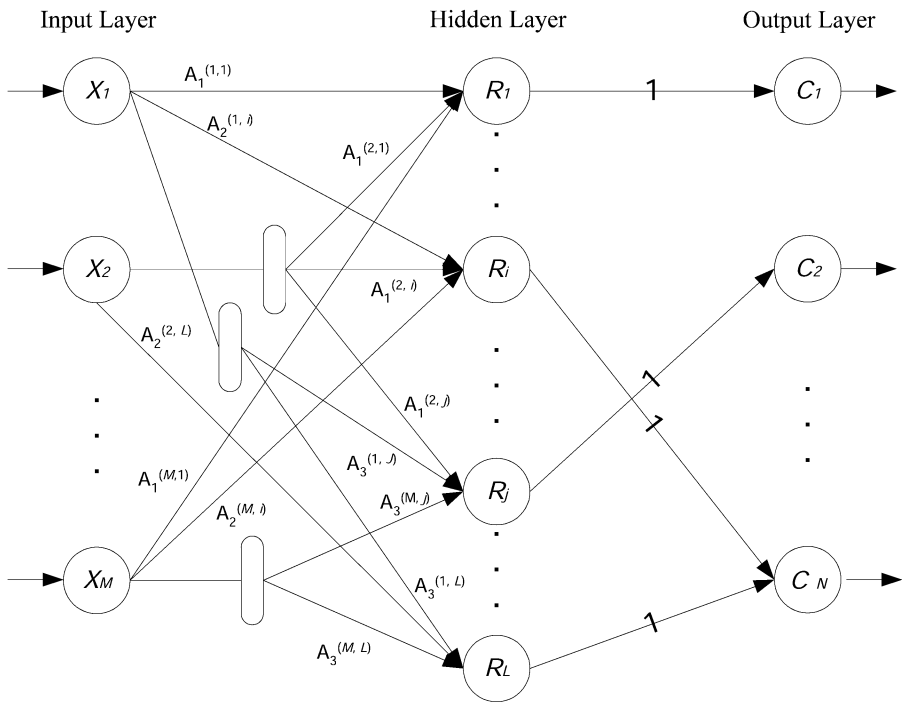

Fig. 3. Architecture of a neuro-fuzzy system (adapted from [34]).

errors cannot be further reduced. Like a backpropagation neural network, the membership functions of fuzzy sets are trained by a backpropagation-like algorithm [42].

C. Feature Selection and Pruning

One of the most challenging tasks in building a classification model is to determine the optimum set of features, which consists of the minimum number of features that realizes at least comparable class separability with the initial set [38]. Reduction in the number of features is necessary to construct a compact classification system and to realize high classification performance. The optimum set of features is commonly determined by feature selection. Chakarborty *et al.* [43] developed a neuro-fuzzy scheme that integrated feature selection and pruning into the neural network scheme. It has been shown that pruning could produce easily interpretable rules and drastically reduce the size of the network without hurting the performance. Four pruning strategies have been implemented in NEFCLASS that support feature selection after the initial model is determined based on potentially useful input features [42]. The strategies focus on testing linguistic terms or fuzzy sets or even variables in the rules, and the variables are removed if doing so improves accuracy, as shown in the following.

a) Delete a linguistic term from the antecedent of a rule if the correlation between the input and output variables suggests that the former is not important.
b) Delete rules that never or very rarely provide the maximum degree of fulfillment for the class given by their consequents.
c) Delete a linguistic term from the antecedent of a rule if the variable is not important for the degree of fulfillment of the rule.
d) Delete a linguistic term from the antecedent of a rule if the term uses fuzzy set with a very large support.

During pruning, the removing of items is followed by consistency checking of the rule base and retraining of the fuzzy sets. If the new classifier is better, it is kept and the pruning steps are repeated. Otherwise, the pruning stops and the previous classifier is recovered. Therefore, the pruning strategies help feature and fuzzy sets selection, reducing the size of the rule base, and increasing the interpretability of learned rules, especially for high-dimensional problems. By applying the neuro-fuzzy approach to DD in interpersonal communication, we can potentially advance our knowledge and improve the performance of DD.

D. Experiment Design

The objective of the experiments was to assess the effectiveness of representation and reasoning under uncertainty in DD by evaluating the performance of a fuzzy DD rule base discovered with the neuro-fuzzy approach. The class labels were deceptive (D) and truthful (T), and a participant can be classified as D, T, or BOTH with a degree of fulfillment ranging from zero to one.

1) Description of Data Sets: Despite the lack of publicly available benchmark deception data sets, largely due to the challenges and practical implications of this line of research, we were able to get access to a diverse set of data owing to past empirical studies. All the data were collected in the United States, and participants were native English speakers. Given the impact of communication modality and task type on cues to deception and DD [6], [19], [44], [45], we selected deception data sets collected from an array of communication settings. The ommunication modalities ranged from face-to-face (FtF), audio, and text-based instant messaging to e-mail. The task types involved both interview and group decision making. These allowed us to

TABLE II
DESCRIPTION OF THE DATA SETS

Datasets	Modalities	Task types	Data sizes*
DSP1	Email	Group decision making	60 (16/44)
DSP3	Text-based Instant messaging (IM)	Group decision making	67 (27/40)
MT0	Text-based Instant messaging	Interview	26 (11/15)
MT1	Face-to-face (FtF)	Interview	31 (12/19)
MT2	Audio	Interview	25 (15/10)

* Data sizes (deceptive/ truthful)

evaluate the neuro-fuzzy approach across different modalities and task types.

The selected data sets and their communication settings are described in Table II. Task scenarios used to elicit deceptive communication were desert survival problem (DSP) [6] and mock theft (MT) [19], respectively. In the DSP task, participants were given descriptions of made-up scenarios in which they were stranded in the middle of a desert. With 15 items at their disposal such as a jack and canvas, they needed to discuss the importance of salvaged items to their survival. Deceivers were induced to deviate from their true belief during the discussion. In the MT task, half of the participants were informed of someone's stealing a wallet from a classroom. Each participant received a standard interview to respond to the interviewer that they did or did not steal the wallet. Deceivers were instructed to persuade the interviewer that they did not steal the wallet, although they actually did, and vice versa.

The messages and speech generated from the above communication contexts were recorded and transcribed into text, if necessary. The unit of analysis was participant. The majority of cues were extracted from verbal behavior, which manifested in the content and language of the text. Lexical diversity and self-references are typical examples of verbal behavior. A small number of cues such as pause and response delay were extracted from nonverbal behavior. For the sake of space, descriptive statistics of the large number of input variables are not listed here.

2) Evaluation Metrics: Accuracy is a common criterion for evaluating a classifier. To reflect possible bias of a DD model, we adopted three measures: deception accuracy, truth accuracy, and overall accuracy. They were defined as the percentage of deceptive, truthful, and all participants that were classified correctly. All the results were based on ten runs of tenfold cross-validations to test the generality of the models on new data. For each run of a tenfold cross-validation, the instances were randomly divided into ten stratified samples. The first sample was used for testing and the other nine samples for building the classifier. Then, the next sample was used for testing and the other nine samples for building the classifier, and so on. At the end of each run, all the cases were used for testing.

To further enhance the validity of the results, all the accuracies were measured at the 95% confidence interval of correct classification per unseen datum. Additionally, standard deviations of classification accuracies and coefficient variations were reported to measure the stability of neuro-fuzzy classifiers.

Fuzzy classifiers are not a replacement of other types of classifiers. Thus, a secondary objective of the evaluation was to assess the relative performance of neuro-fuzzy techniques by comparing their results with those of traditional machine-learning techniques including neural networks and decision trees.

3) Procedure: Following the general procedure for DD and neuro-fuzzy learning [38], as discussed in Sections II and III, we built the DD models. Additional details about feature identification and parameter estimation are described as follows.

We started with all the features from the original data sets. Most features were extracted from the text via natural language processing and postprocessing (see [6] for more details). For example, self-reference was identified based on the parts of speech; and pleasantness was recognized by combining morphological analysis and specialized lexical resources. Some cues were derived from accessory information of messages. For example, pause was computed as the difference between the sending time of two consecutive messages. The features were pruned and their membership functions tuned to optimize the classification accuracy during feature optimization.

An initial step in building a neuro-fuzzy system involves estimating initial model parameters and constraints. In particular, granularity of fuzzy sets for each cue was set to three or five depending on the value distribution of the cue; and the relative order and symmetry of fuzzy sets was preserved while allowing them to overlap in order to get interpretable rules. A triangular membership function was used because it was most widely used for representing uncertainty due to fuzziness. Maximum aggregation function selected the consequent of a rule with the highest degree of fulfillment when a crisp decision was needed. For the neural network, the learning rate was set to 0.01 and maximum number of epochs was 100.

IV. RESULTS AND DISCUSSION

In this section, we first evaluate the performance of the neuro-fuzzy DD models. Then, we discuss the findings of this study and their implications.

A. Results

The results of neuro-fuzzy models were reported in Tables III and IV. The unit of the errors for confidence intervals (CIs) was a single run of tenfold cross-validation in Table III and was each of the ten runs of tenfold cross-validation in Table IV. As shown in Table III, the 95% CI of the overall accuracies of optimal neuro-fuzzy classifiers varied from [56, 75] to [67, 99]. Specifically, deception accuracies range from 18.5% to 93.3% and truth accuracies from 30% to 97.7%. The above results were produced by pruned classifiers, which were improved greatly over those of classifiers without pruning. For example, in a single run of cross-validation, pruning improved the 95% CI of neuro-fuzzy models for DSP1 and MT1 from [0, 12] to [67, 86] and from 0 to [67, 99], respectively.

It is shown from Table IV that the ranges of means and standard deviations of the overall classification accuracies are [60.7%, 81.3%] and [2.1%, 9.6%], respectively. In addition, coefficients of variation range from 0.035 to 0.128. Both

TABLE III
RESULTS OF OPTIMAL NEURO-FUZZY MODELS FROM A SINGLE RUN OF TENFOLD CROSS-VALIDATION

Datasets	# of rules	# of features	Classification Accuracies (%)					
			Training[1]			Testing		
			Overall	Deception	Truth	95% C.I. [2]	Deception	Truth
DSP1	12	12	85	56.3	95.5	[67, 86]	31.3	97.7
DSP3	10	4	67.2	25.9	95	[56, 75]	18.5	85.0
MT0	6	3	69.2	63.3	73.3	[64, 99]	81.8	86.7
MT1	7	8	80.6	66.7	89.5	[67, 99]	66.7	94.7
MT2	8	3	80	93.3	60	[60, 90]	93.3	30.0

[1]…The final model based on a tenfold cross validation using all the data as both training data and testing data
[2]…95% confidence interval for percentage of correct classification per pattern for unseen data using a tenfold cross validation

TABLE IV
DESCRIPTIVE STATISTICS OF THE OVERALL PERFORMANCE OF PRUNED NEURO-FUZZY MODELS

Datasets	Means (%)	Standard deviations (%)	Coefficients of variation	95% C.I. (%)
DSP1	81.3	3.46	0.043	[78, 84]
DSP3	60.7	2.12	0.035	[59, 63]
MT0	75.27	9.61	0.128	[68, 83]
MT1	80.2	5.38	0.067	[76, 84]
MT2	66.2	5.27	0.080	[62, 70]

standard deviations and coefficient variations for the DSP data sets are lower than those for the MT data sets, suggesting that neuro-fuzzy models are more stable for DD with decision-making tasks than interview tasks.

Traditional machine-learning techniques such as neural networks (NNs) and decision trees (DTs) have been applied to automatic DD [2], [3]. Therefore, we also compared the performance of neuro-fuzzy (NF) models with those of NN [46] and DT models [47]. A paired t test based on random split has been shown to have a high probability of Type I error in some situations while comparing supervised classification learning algorithms [48]. Although the violation of the assumption of independent test sets is relieved, k-fold cross-validated paired t test still suffers from the problem of the overlapping training sets. It has been shown that ten-run tenfold cross-validation has better replicability in addition to an appropriate Type I error and low Type II error [49]. Therefore, we ran tenfold cross-validation ten times along with corrected repeated k-fold cross-validation tests [49], [50] on five different data sets to compare the overall performance of the three machine-learning techniques [see (1)]. The results are reported in Fig. 4 and Table V.

$$t = \frac{\frac{1}{k*r}\sum_{i=1}^{k}\sum_{j=1}^{r} x_{ij}}{\sqrt{\left(\frac{1}{k*r}+\frac{n_2}{n_1}\right)\hat{\sigma}^2}} \quad (1)$$

where n_1 is the number of instances used for training, n_2 is the number of instances used for testing, k and r are the number of folds and runs for cross-validation, respectively, $x_{ij} = a_{ij} - b_{ij}\ (1 \leq i \leq k, 1 \leq j \leq r)$,, and $\hat{\sigma}^2 = 1/(k \cdot r - 1)\sum_{i=1}^{k}\sum_{j=1}^{r}(x_{ij} - m)^2$ and $m = 1/(k \cdot r)\sum_{i=1}^{k}\sum_{j=1}^{r} x_{ij}$ are the estimates for the variance and the mean, respectively.

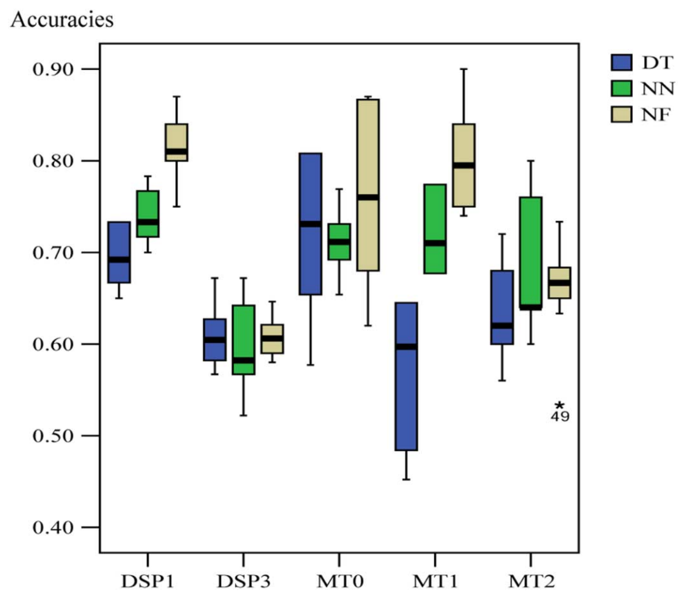

Fig. 4. Comparison of three types of DD models.

TABLE V
T-STATISTICS FOR PERFORMANCE COMPARISONS BETWEEN THREE TECHNIQUES

Datasets	NF vs. NN	NF vs. DT
DSP1	0.651129	0.615357
DSP3	0.055872	-0.00342
MT0	0.106346	0.06441
MT1	0.370056	0.915489
MT2	-0.04049	0.112266

-: the performance of NF is lower than the other approach in comparison

B. Discussion

The results showed that the performance of the neuro-fuzzy classifiers was satisfactory in comparison with other types of classifiers previously used in DD. In particular, neuro-fuzzy classifiers achieved higher accuracies than neural networks and decision trees in eight out of ten pairwise comparisons even though the improvements were not statistically significant. This paper provides the first evidence showing that neuro-fuzzy techniques are a good alternative for developing DD models.

Given that the data sets used in this paper were collected from two types of tasks (i.e., DSP and MT) and three types of modalities (i.e., text, audio, and FtF), it would be interesting to compare the performance of neuro-fuzzy models across different settings. Using the means reported in Table IV, we conducted

ANOVA and Tamhane's T2—a conservative pairwise comparisons test based on t test. It was found that, for the DSP task, the neuro-fuzzy technique performed better on DSP1 than DSP3 ($p < 0.001$). This suggests that it is easier to detect deception from asynchronous communication (i.e., e-mail) than synchronous communication (i.e., instant messaging). For the MT task, the performance of neuro-fuzzy models was better in FtF than in the audio modality ($p < 0.001$). FtF is a richer medium that offers more cues than the audio modality [28], [51]. FtF modality also induces a higher level of arousal in deceivers than computer-mediated communication [2]. As the number of incoming and outgoing cues increase, fuzzy methods seem to become a better fit for DD. This is partly because the increase in the number of available cues increases the chance for inconsistency or contradiction between different cues, elevating the amount of uncertainty in DD. The pairwise multiple comparisons also revealed that, given the same modality (i.e., instant messaging), the neuro-fuzzy technique performes better for the interview task than the decision-making task ($p < 0.01$). A comparison of accuracy gains in the text modality between the two types of tasks shows that the gain of neuro-fuzzy models was greater for MT0 than DSP3. Compared with the interview task, the decision-making task has no obvious right answer [52] and introduces situations with a higher level of equivocality [51], which further increases the uncertainty in DD. Fuzzy theory and fuzzy logic closely match the cognitive model of human decision making under uncertainty. Therefore, fuzzy models developed in this paper provide a promising approach to representing and reasoning uncertainty in DD.

It is shown from the comparison of training and test results in Table III that the performance of the neuro-fuzzy model does not degrade seriously after switching from training to holdout data. This trend is the opposite of what some traditional machine-learning models show [2]. This finding suggests that the neuro-fuzzy models are not overtrained and hence generalizable. The poor generalization of traditional approaches could be attributable to their lack of means to handle uncertainty due to greater variance and imprecision in cues to deception and detection outcomes.

Pruning has been found to reduce both the number of rules from [25, 57] to [6, 12] and the number of selected features from [29, 36] to [3, 12]. For example, the ten rules learned for DSP3 account for only 0.02% of the total number of possible rules. The significant reductions demonstrate the effectiveness of neuro-fuzzy learning and pruning strategies. A comparison of the results before and after pruning reveals that the pruning significantly improved the confidence intervals of classification accuracies and reduced the number of errors across all the data sets. Pruning proved to be an extremely powerful method for increasing the performance and interpretability and for improving the generality of classification models.

A comparison of deception accuracies and truth accuracies in Table III showed that neuro-fuzzy models generally identify truth better than deception. Thus, neuro-fuzzy models appear to have deception bias. Since the number of deceptive and truthful samples in the data sets was more or less balanced, we may not be able to generalize the above observation to other highly skewed data sets.

Fuzzy classification normally does not offer a clear classification decision. Instead, it generates a vector of real values in the range of [0, 1]. In particular, the neuro-fuzzy approach allows an instance to be classified as both deceptive and truthful with different degrees of fulfillment. For example, an instance was assigned degrees of fulfillment of 0.547 for deceptive and 0.373 for truthful respectively. When a "maximum" aggregation function is applied, the above instance would be classified as deceptive. In case the degrees of fulfillments for deceptive and truthful classes are tied, the instance would be unclassified. Such information can be extremely valuable for human DD.

This representation of imprecise cues to deception with fuzzy sets can be used to handle sophisticated cases. Depending on a person's tolerance of deception and the cost associated with missing or misidentifying a deceptive instance, he or she can take advantage of the additional information in making a detection decision. As a result, the neuro-fuzzy models not only help humans in DD by providing useful information but also enhance the human's ability to detect deception.

The parameter tuning in a developing neuro-fuzzy system remains time-consuming. For example, the number of fuzzy sets for each feature is determined based on its data distribution. Such a decision could impact whether the feature will be selected into the final model. Thus, neuro-fuzzy training is an interactive process. Solutions that can streamline the above process while preserving the strengths of a rule-based fuzzy system are highly desired. Additionally, a neuro-fuzzy system allows for incorporating domain experts' heuristic knowledge into the training process such as feature representation and rule pruning. However, it is not an easy task to translate heuristic knowledge into the fuzzy set framework. A heuristic tuning process will be required in order to get an optimized system.

C. Discovered Knowledge

One major advantage of the neuro-fuzzy approach is the ability to generate comprehensible rules. These rules become valuable knowledge for human and automatic DD and are discussed in detail in this section. The final sets of rules discovered from this paper are listed in Table VI. Features (cues to deception) contained in the fuzzy rules are listed in Table VII. Three sample cues from the output of DSP1 and their fuzzy sets are presented in Fig. 5.

The selected features, as shown in Table VII, provide new evidence to support recent findings on cues to deception in computer-mediated communication as well as previous findings in face-to-face communication. For example, deceivers in DSP1 were found to use more modal verbs, group references, misspelled words, modifier verbs, and affect and to have lower word diversity and pausality [6]. Pleasantness, positive pleasantness, imagery, and positive imagery were identified as useful cues to deception in an earlier study of DSP1 [2]. These cues can be interpreted from the communication and psychology perspectives. A lower level of lexical diversity in deceivers could result from high cognitive effort and arousal involved in deception. Higher productivity enhances the persuasiveness of deceivers' messages, which in turn help them achieve their deceptive goals.

TABLE VI
RULES DISCOVERED BY THE NEURO-FUZZY MODELS

Data sets	Fuzzy rules (D: deceptive; T: truthful; &: and; S: small; L: large; M: medium; =>: if-then)
DSP1	modal verb is S & affect is S & group reference is L => D.
	model verb is L & affect is L & content diversity is S & group reference is L => D.
	modal verb is S & affect is S & group reference is M => T.
	modal verb is L & group reference is S => T.
	modal verb is M & affect is S & group reference is M & pleasantness is S & positive pleasantness is S => T.
	verb is very S & modifier is very S & modal verb is S & affect is S & content diversity is L & group reference is S => T.
	pausality is S & modifier is very L & model verb is L & content diversity is L & group reference is M & pleasantness is S & imagery is S & positive pleasantness is M & positive imagery is S & negative imagery is L => D.
	pausality is S & modifier is S & modal verb is M & effect is S & content diversity is M & group reference is M & pleasantness is S & imagery is S & positive pleasantness is L & positive imagery is S & negative imagery is L => D.
	pausality is S & verb is very S & modifier is very S & modal verb is M & content diversity is S & group reference is S & pleasantness is L & imagery is S & positive pleasantness is L & positive imagery is S & negative imagery is L => D.
	pausality is S & modifier is M & modal verb is M & content diversity is M & group reference is M & pleasantness is S & imagery is S & positive pleasantness is M & positive imagery is S & negative imagery is L => T.
	pausality is S & verb is L & modifier is L & modal verb is M & content diversity is S & group reference is M & pleasantness is S & imagery is S & positive pleasantness is L & positive imagery is S & negative imagery is L => T.
	pausality is S & verb is very S & modifier is very S & modal verb is S & affect is L & content diversity is L & group reference is S & pleasantness is S & imagery is L & positive pleasantness is S & positive imagery is L & negative imagery is S => T.
DSP3	#message is S & pause is extremely S & delay is very L => D.
	#message is very L & pause is extremely S & delay is S => T.
	#message is very S & pause is extremely S & delay is S => T.
	#message is S & pause is extremely S & delay is M => T.
	#message is very S & pause is extremely L & delay is very S & lexical diversity is L => T.
	#message is very S & pause is extremely S & delay is very S => T.
	#message is S & pause is extremely S & delay is very S => T.
	#message is S & pause is extremely S & delay is S => T.
	#message is M & pause is extremely S & delay is S => D.
	#message is S & pause is extremely S & delay is L & lexical diversity is L => T.
MT0	personal reference is L & activation is M => D.
	personal reference is S & activation is L & positive activation is L => D.
	personal reference is L & activation is L & positive activation is L => D.
	personal reference is S & activation is S & positive activation is S => T.
	personal reference is M & activation is L & positive activation is L => D.
	personal reference is L & activation is S & positive activation is S => T.
MT1	sensory is S & positive activation is L & negative pleasantness is M => D.
	negative pleasantness is S => T.
	positive activation is S & negative pleasantness is L => T.
	pausality is S & sensory is M & specificity is L & sentence length is L & redundancy is L & activation is M & positive activation is M & negative pleasantness is M => D.
	pausality is L & sensory is M & temporal non-immediacy is L & spatial immediacy is S & specificity is L & sentence length is L & redundancy is L & activation is M & positive activation is L & negative pleasantness is M => D.
	pausality is S & affect is L & sensory is S & spatial immediacy is S & specificity is S & sentence length is S & redundancy is S & activation is M & positive activation is M & negative pleasantness is M => D.
	pausality is S & sensory is M & temporal non-immediacy is S & specificity is M & sentence length is S & redundancy is S & activation is M & positive activation is S & negative pleasantness is M => T.
MT2	spatial non-immediacy is M & self reference is L & highly positive activation is S => D.
	spatial non-immediacy is S & self reference is L & highly positive activation is L => D.
	spatial non-immediacy is S & self reference is S & highly positive activation is S => D.
	spatial non-immediacy is S & self reference is M & highly positive activation is S => D.
	spatial non-immediacy is S & self reference is S & highly positive activation is L => T.
	spatial non-immediacy is L & self reference is L & highly positive activation is L => T.
	spatial non-immediacy is L & self reference is S & highly positive activation is M => T
	spatial non-immediacy is M & self reference is S & other reference is M & highly positive activation is S => T.

This is especially true in distributed computer-mediated communication, where more planning time is allowed. The negative experience from deception may be transformed into more words reflecting low pleasantness in deceivers' messages, higher level of dynamics of emotional state (i.e., activation), and less frequent usage of words that provide a clear mental picture (i.e., imagery) [2].

Lexical diversity and participation (i.e., pause and delay) were also found to be significant cues to deception in a previous study of DSP3 [5]. In particular, deceivers took shorter pauses and had lower lexical diversity than truth-tellers. Our findings on cues were also consistent with other studies of the MT data [3], [53]. For example, in comparison to truth-tellers, deceivers in the text modality used more personal references and spatial nonimmediacy and less sentence length, pausality, and modal verbs; deceivers in the face-to-face showed more affect; and deceivers in the audio modality showed more temporal nonimmediacy and less word length and pausality. Some new cues selected by this study were attributable to cognitive, emotional, and psychological arousals experienced by intentional deceivers during deception [1]. For example, deceivers showed higher nonimmediacy, negative pleasantness, and imagery and

TABLE VII
CUES EXTRACTED FROM THE FUZZY RULES

Datasets	Selected Cues (In the descending order of significance)
DSP1	modal verb, group reference, lexical diversity, misspelled word, affect, modifier, pleasantness, positive pleasantness, pausality, imagery, positive imagery, negative imagery, verb
DSP3	#messages, pause, delay, lexical diversity
MT0	personal reference, activation, positive activation
MT1	redundancy, imagery, positive activation, temporal non-immediacy, spatial immediacy, sensory, activation, negative pleasantness
MT2	spatial non-immediacy, self reference, highly positive activation

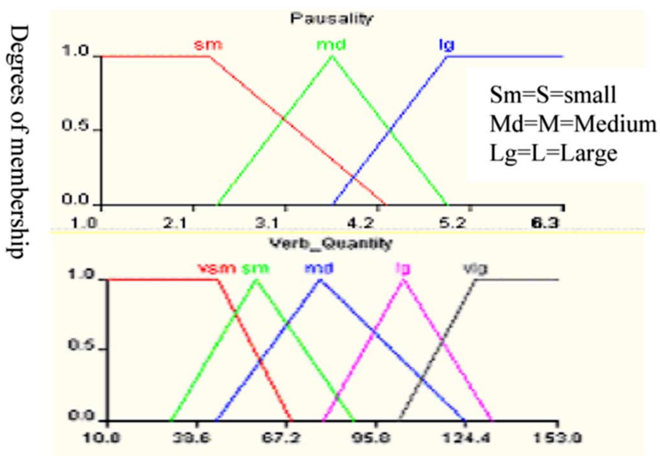

Fig. 5. Fuzzy set representation of two cues.

lower self-references, sensory perception, and immediacy. In addition, high redundancy in deceptive messages helped reduce the cognitive complexity induced by deception.

It is noted from the fuzzy rules in Table VI that a fuzzy linguistic term may not always be associated with the deceptive or the truthful class. For example, for MT0, *personal reference is small* was included in a rule for predicting deceptive and in another rule for predicting truthful. Although the strengths of the rules being fired are different, they suggest that deception is likely to be an interplay of multiple behavioral cues rather than a single-factor display. Accordingly, a holistic view of multiple cues is necessary to reduce possible wrong detection. The phenomena highlight the complexity and uncertainty of DD.

The results on DSP3 were worse than other data sets across all the machine-learning techniques. We found several alternative explanations for it. First, compared with e-mails in DSP1, instant messages) in DSP3 were more spontaneous and less formal. The informal language style in IM may have reduced the effectiveness of linguistics cues. In addition, the majority of cues in DSP3 were nonverbal but those in DSP1 were linguistics-based cues. Thus, the results from DSP3 may be improved if a broad range of linguistic cues were incorporated. Secondly, participants had more interactions during IM than in e-mail communication and more frequent and direct interaction during group decision making than during interview. As a result, there may be a fair amount of reciprocity of behavior displayed [54] between deceivers and truth-tellers in DSP3, making it difficult to tell the deceiver from the truth-teller [2]. Thirdly, the nature of the relationship between two parties involved in ongoing interactions is often determined by the purpose and formality of the context and the time the group has been together [11]. The reciprocity is higher in intimate relationships [55] and possibly decision making, which further increases the difficulty of DD.

This paper demonstrates that neuro-fuzzy classifiers, which exploit uncertainty in DD, are able to not only predict deception as accurately as the best traditional machine-learning techniques but also provide degrees of support in predicting deception. Moreover, fuzzy rules learned by the neuro-fuzzy classifiers exhibit good interpretability by conforming to humans' heuristic knowledge applied to DD.

D. Implications

The findings of the study reported in this paper have theoretical, methodological, and practical implications to DD and fuzzy systems research. The study advances fuzzy systems research by extending fuzzy theory and neuro-fuzzy systems to handle uncertainty due to vagueness and imprecision to a new application domain—DD. The theoretical model of uncertainty in DD can improve our understanding of deception and lend itself to a theory of DD. The finding that fuzzy theory and fuzzy logic become more effective for DD under a higher level of uncertainty highlights the importance of matching methods to the characteristics of the problem under investigation.

The computational model of representation and reasoning under uncertainty lays the groundwork for handling uncertainty in DD. The promising and stable performance of the neuro-fuzzy classifier in DD across multiple modalities and task types motivates further work to advance techniques for DD. In addition, empowering the learning process by incorporating humans' domain knowledge put neuro-fuzzy classifiers to an advantageous position. Thus, the performance of neuro-fuzzy models would improve as our knowledge about DD advances.

Focusing on uncertainty itself puts the study of DD in a more realistic setting. The fuzzy models can be used not only to detect deception automatically but also to serve as a decision aid to human DD in the real world. This is enabled by the fuzzy rules that are not only interpretable but also associated with strengths and degrees of confidence in making deceptive and truthful predictions. Moreover, fuzzy rules and cues to deception discovered in this paper alleviate the knowledge bottleneck problem in DD, especially in computer-mediated communication. They can be applied to educate and train people in detecting deception. Since fuzzy rules fall between natural language expression and formal mathematical representation [56], training people on cues to deception could become much easier. Furthermore, the models can be applied to deter and prevent future deception by reducing deceivers' propensity to commit deception [57].

V. Conclusion and Future Work

To the best of our knowledge, this is the first study to examine uncertainty in DD from both theoretical and technical perspectives. Based on a systematic analysis of uncertainty in DD, we proposed a computational model for representing and reasoning uncertainty in DD. Fuzzy theory and fuzzy logic were used to address uncertainty due to imprecision and vagueness in DD. Neuro-fuzzy systems were developed for DD by automatically learning rules from the data.

The findings of this paper suggest that, in terms of performance, reliability, and interpretability, techniques from the soft computing paradigm (e.g., neuro-fuzzy models) can be a promising alternative to some traditional machine-learning techniques for DD. The neuro-fuzzy technique even showed a better performance in DD involving a higher level of uncertainty.

DD is a challenging task. Given the ever-present uncertainty in DD, especially in high-stake situations, it is highly desired to exploit uncertainty in assisting humans to make detection decisions. The DD models introduced here can be extended to construct and build DD systems in other types of communication contexts. For users, making a detection decision is not the end, and what needs to be followed is to develop and take a course of rectification actions [58]. It will be an interesting and practical problem to develop post-detection decision support for end-users.

There are basically six avenues along which this research should be continued. First, models can be developed to address other types of uncertainty in DD such as uncertainty due to randomness. Secondly, given the importance of parameter settings in neuro-fuzzy training and optimization, methods for improving parameters estimation and tuning are highly desired. Some preliminary evidence has shown the potential of genetic algorithms in fuzzy systems development [59], [60]. Thirdly, the usability of fuzzy rules can be enhanced. There could be multiple rules matching one instance. Although only the rule with the largest strength is fired, reduction in the overlap between the linguistic terms of different rules and increase in the association between linguistic terms and the deceptive (or truthful) class may further improve the interpretability of the rules. Fourthly, comparing the parameters of the model-derived fuzzy sets against the parameters of the human-created fuzzy sets would provide a better picture of how well the neuro-fuzzy approach works for DD. It requires information about participants' degrees of membership in deceptive and truthful classes, which will be investigated in a follow-up study. Fifthly, user studies should be carried out to test the effectiveness of the fuzzy rules and significance of the fuzzy sets in human DD. To this end, a decision support system for human DD can be built and evaluated. Sixthly, a fusion or ensemble approach to DD could help improve the confidence in an overall DD decision by considering the output of multiple DD classification models.

Acknowledgment

The authors thank Prof. J. Burgoon, Prof. J. Nunamaker, and T. Qin of CMI at the University of Arizona for sharing some of their data with them. They would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

[1] D. B. Buller and J. K. Burgoon, "Interpersonal deception theory," *Commun. Theory*, vol. 6, pp. 203–242, 1996.
[2] L. Zhou, D. Twitchell, J. Burgoon, T. Qin, and J. Nunamaker, "A comparison of classification methods for predicting deception in computer-mediated communication," *J. Manage. Inform. Syst.*, vol. 20, pp. 139–165, 2004.
[3] J. K. Burgoon, J. P. Blair, T. Qin, and J. F. Nunamaker, "Detecting deception through linguistic analysis," presented at the 1st NSF/NIJ Symp. Intell. Security Informatics , Tucson, AZ, 2003, LNCS 2665, unpublished.
[4] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception," *Psychol. Bull.*, vol. 129, pp. 74–112, 2003.
[5] L. Zhou, "An empirical investigation of deception behavior in instant messaging," *IEEE Trans. Prof. Commun.*, vol. 48, pp. 147–160, 2005.
[6] L. Zhou, J. K. Burgoon, J. F. Nunamaker, and D. Twitchell, "Automated linguistics based cues for detecting deception in text-based asynchronous computer-mediated communication: An empirical investigation," *Group Decision Negot.*, vol. 13, pp. 81–106, 2004.
[7] D. P. Twitchell and D. P. Biros, "Automated determination of the veracity of interview statements from people of interest to an operational security force," presented at the 39th Hawaii Int. Conf. Syst. Sci., Kauai, HI, 2006, unpublished.
[8] A. Vrij and S. Mann, "Detecting deception: The benefit of looking at a combination of behavioral, auditory, and speech content related cues in a systematic manner," *Group Decision Negot.*, vol. 13, pp. 61–79, 2004.
[9] J. K. Burgoon, N. Miczo, and L. A. Miczo, "Adaptation during deceptive interactions: Testing the effects of time and partner communication style," presented at the Nat. Commun. Assoc. Conv., Atlanta, GA, 2001, unpublished.
[10] C. H. White and J. K. Burgoon, "Adaptation and communicative design: Patterns of interaction in truthful and deceptive conversation," *Human Commun. Res.*, vol. 27, pp. 9–37, 2001.
[11] L. Zhou, J. K. Burgoon, D. Zhang, and J. F. Nunamaker, "Language dominance in interpersonal deception via computer-mediated communication," *Comput. Human Behav.*, vol. 20, pp. 381–402, 2004.
[12] C. Joslyn and L. Rocha, "Towards a formal taxonomy of hybrid uncertainty representations," *Inform. Sci.*, vol. 110, pp. 255–277, 1998.
[13] L. A. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338–353, 1965.
[14] J.-Y. Mao and I. Benbasat, "The use of explanations in knowledge-based systems: Cognitive perspectives and a process-tracing analysis," *J. Manage. Inform. Syst.*, vol. 17, pp. 153–179, 2000.
[15] Z. Chen, *Data Mining and Uncertain Reasoning: An Integrated Approach*. New York: Wiley, 2001.
[16] R. Weber, "Customer segmentation for banks and insurance groups with fuzzy clustering techniques," in *Fuzzy Logic*, J. Baldwin, Ed. New York: Wiley, 1996, pp. 187–196.
[17] G. Johnson and E. Santos, "Detecting deception in intelligent systems I: Activation of deception detection tactics," presented at the Adv. Artif. Intell. 17th Conf. Can. Soc. Comp. Studies Intell. , Ontario, ON, Canada, 2004, LNCS, unpublished.
[18] B. Yu and M. P. Singh, "Detecting deception in reputation management," presented at the 2nd Int. Joint Conf. Auton. Agents Multi-Agent Syst., Melbourne, Australia, 2003, unpublished.
[19] J. K. Burgoon, J. P. Blair, and R. E. Strom, "Heuristics and modalities in determining truth versus deception," presented at the 38th Hawaii Int. Conf. Syst. Sci., Big Island, HI, 2005, unpublished.
[20] B. M. DePaulo, J. T. Stone, and G. D. Lassiter, "Deceiving and detecting deceit," in *The Self and Social Life*, B. R. Schlenker, Ed. New York: McGraw-Hill, 1985, pp. 323–370, unpublished.
[21] T. O. Boucher, "Reliability, validity, and imprecision in fuzzy multicriteria decision-making," *IEEE Trans. Syst., Man, Cybern. C*, vol. 32, pp. 190–202, 2002.
[22] N. S. Iyer, "A family of dominance rules for multiattribute decision making under uncertainty," *IEEE Trans. Syst., Man, Cybern. A*, vol. 33, pp. 441–450, 2003.
[23] H. C. Bahl and R. G. Hunt, "Decision-making theory and DSS design," *SIGMIS Database*, vol. 15, pp. 10–14, 1984.
[24] L. Zhou and A. Zenebe, "Modeling and handling uncertainty in deception detection," presented at the 38th Hawaii Int. Conf. Syst. Sci., Big Island, HI, 2005, unpublished.

[25] G. J. Klir and T. A. Folger, "," in *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, NJ: Prentice-Hall, 1988, unpublished.
[26] P. Smets, "Theories of uncertainty," in *Handbook of Fuzzy Computation*, E. Ruspini, P. P. Bonissone, and W. Pedrycz, Eds. Philadelphia, PA: Institute of Physics, 1999, unpublished.
[27] M. Smithson, *Ignorance and Uncertainty: Emerging Paradigms*. New York: Springer-Verlag, 1989.
[28] R. L. Daft and R. H. Lengel, "Organizational information requirements, media richness and structural design," *Manage. Sci.*, vol. 32, pp. 554–571, 1986.
[29] C. Olaru and L. Wehenkel, "A complete fuzzy decision tree technique," *Fuzzy Sets Syst.*, vol. 138, pp. 221–254, 2003.
[30] T. G. Dietterich and E. B. Kong, Machine learning bias, statistical bias, and statistical variance of decision tree algorithms Dept. of Comput. Sci., Oregon State Univ., Tech. Rep., 1995.
[31] L. L. Kuncheva, *Fuzzy Classifier Design*. New York: Springer-Verlag, 2000, vol. 49.
[32] J. H. Friedman, Local learning based on recursive covering Dept. of Statistics, Stanford Univ., Technical Rep., Aug. 1996.
[33] S. P. Curram and J. Mingers, "Neural networks, decision tree induction and discriminant analysis: An empirical comparison," *J. Oper. Res. Soc.*, vol. 45, pp. 440–450, 1994.
[34] D. Nauck, F. Klawonn, and R. Kruse, *Foundations of Neuro-Fuzzy Systems*. Chichester, U.K.: Wiley, 1997.
[35] Q. Chen, "Pattern recognition and classification in user modeling: A neural network approach," in *Information Systems*. Baltimore, MD: Univ. of Maryland Baltimore County, 1995, p. 174.
[36] J. Pearl, "Decision making under uncertainty," *ACM Comput. Surv.*, vol. 28, pp. 89–92, 1996.
[37] L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Commun. ACM*, vol. 37, pp. 77–84, 1994.
[38] S. Abe, *Pattern Classification: Neuro-Fuzzy Methods and Their Comparison*. London, U.K.: Springer-Verlag, 2001.
[39] M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools," *SIGKDD Explor. Newsl.*, vol. 1, pp. 20–33, 1999.
[40] E. Cox, *The Fuzzy Systems Handbook*, 2nd ed. New York: AP Professional, 1999.
[41] T. Bilgiç and I. B. Turksen, "Measurement of membership functions: Theoretical and empirical work," in *Fundamentals of Fuzzy Sets, Vol. 1, Handbook of Fuzzy Sets and Systems*, D. Dubois and H. Prade, Eds. Boston, MA: Kluwer, 1999, pp. 195–232.
[42] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," *Artif. Intell. Med.*, vol. 16, pp. 149–169, 1999.
[43] D. Chakraborty and N. R. Pal, "A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification," *IEEE Trans. Neural Netw.*, vol. 15, pp. 110–123, 2004.
[44] B. M. DePaulo, M. Zuckerman, and R. Rosenthal, "Detecting deception: Modality effects," in *Review of Personality and Social Psychology*, L. Wheeler, Ed. Beverly Hills, CA: Sage, 1980, pp. 125–162.
[45] G. A. Giordano and J. F. George, "Task complexity and deception detection in a collaborative group setting," presented at the 38th Hawaii Int. Conf. Syst. Sci., Big Island, HI, 2005, unpublished.
[46] D. E. Rumerlhart and J. L. McClelland, *Parallel and Distributed Processing: Exploration in the Microstructure of Cognition*. Cambridge, MA: MIT Press, 1986.
[47] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.
[48] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, pp. 1895–1923, 1998.
[49] R. Bouckaert and E. Frank, "Evaluating the replicability of significance tests for comparing learning algorithms," presented at the Pacific-Asia Conf. Knowl. Discovery Data Mining, Sydney, Australia, 2004, unpublished.
[50] C. Nadeau and Y. Bengio, "Inference for the generalization error," *Machine Learn.*, vol. 52, pp. 239–281, 2003.
[51] A. R. Dennis and S. T. Kinney, "Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality," *Inform. Syst. Res.*, vol. 9, pp. 256–274, 1998.
[52] J. E. McGrath, *Groups: Interaction and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[53] T. Qin, J. Burgoon, J. P. Blair, and J. Nunamaker, "Modality effects in deception detection and applications in automatic deception detection," presented at the 38th Hawaii Int. Conf. Syst. Sci., Big Island, HI, 2005, unpublished.
[54] J. K. Burgoon, L. Dillman, and L. A. Stern, "Adaptation in dyadic interaction: Defining and operationalizing patterns of reciprocity and compensation," *Commun. Theory*, vol. 3, pp. 295–316, 1993.
[55] M. E. Roloff, "Communication and reciprocity within intimate relationships," in *Interpersonal Processes: New Directions in Communication Research*, M. E. Roloff and G. R. Miller, Eds. Beverly Hills, CA: Sage, 1987, pp. 11–38.
[56] R. R. Yager, "Fuzzy modeling for intelligent decision making under uncertainty," *IEEE Trans. Syst., Man, Cybern.*, vol. 30, pp. 60–70, 2000.
[57] D. W. Straub and R. J. Welke, "Coping with systems risk: Security models for management decision-making," *MIS Quart.*, vol. 22, pp. 441–469, 1998.
[58] A. Vyas and L. Zhou, "On detecting deception in agent societies," presented at the IEEE/WIC/ACM Int. Joint Conf. Web Intelligence Intelligent Agent Technol., , France, 2005, Compiegne Univ. of Technol., unpublished.
[59] Y.-S. Chen and C. Shahabi, "Automatically improving the accuracy of user profiles with genetic algorithm," in *Proc. IASTED Int. Conf. Artif. Intell. Soft Comput.*, Cancun, Mexico, 2001.
[60] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases* , ser. Advances in Fuzzy Systems—Applications and Theory. Singapore: World Scientific, 2001, vol. 19.

Lina Zhou (M'99) received the Ph.D. degree in computer science from Peking University, Beijing, China, in 1998.

From 1999 to 2002, she was with University of Arizona, Tucson, as a Research Scientist. She is currently an Assistant Professor of information systems with the University of Maryland Baltimore County. Her research interests are in text mining and machine learning with applications to deception detection, speech, and Web-based information systems problems. She has authored or coauthored more than 25 papers in peer-reviewed professional journals.

Dr. Zhou is a member of ACM and AIS and a Board Member of the Special Interest Group on Semantic Web and Information Systems.

Azene Zenebe received the B.Sc. degree in statistics and M.Sc. degree in information science from Addis Ababa University, Ethiopia, in 1990 and 1996, respectively, and the M.Sc. and Ph.D. degrees in information systems from the University of Maryland Baltimore County in 2002 and 2005, respectively.

He is currently an Assistant Professor of Management Information Systems with Bowie State University, Bowie, MD. His main interests include uncertainty modeling, fuzzy systems, knowledge representation and inference, personalized information systems, machine learning, knowledge discovery, and data mining.