Dempster-Shafer Theoretic Learning of Indirect Speech Act Comprehension Norms

Ruchen Wen, Mohammed Aun Siddiqui and **Tom Williams**
MIRRORLab
Department of Computer Science
Colorado School of Mines
{rwen,twilliams}@mines.edu

Abstract

For robots to successfully operate as members of human-robot teams, it is crucial for robots to correctly understand the intentions of their human teammates. This task is particularly difficult due to human sociocultural norms: for reasons of social courtesy (e.g., politeness), people rarely express their intentions directly, instead typically employing polite utterance forms such as Indirect Speech Acts (ISAs). It is thus critical for robots to be capable of inferring the intentions behind their teammates' utterances based on both their interaction context (including, e.g., social roles) and their knowledge of the sociocultural norms that are applicable within that context. This work builds off of previous research on understanding and generation of ISAs using Dempster-Shafer Theoretic Uncertain Logic, by showing how other recent work in Dempster-Shafer Theoretic rule learning can be used to learn appropriate uncertainty intervals for robots' representations of sociocultural politeness norms.

Introduction and Motivation

For robots to successfully operate as members of human-robot teams, it is crucial for robots to correctly understand the intentions of their human teammates. This not only requires traditional language understanding components such as speech recognition and syntactic and semantic analysis, but also *pragmatic* analysis components for performing deeper analysis with respect to the robot's environmental and social context. This is especially important in contexts that have strong sociocultural norms, conventions, and contracts, since, in such contexts, humans typically phrase their language in the form of *Indirect Speech Acts (ISAs)* (Searle 1975), in which the speech act's literal meaning does not match its intended meaning. For example, in a restaurant context, even though servers are in some sense functioning as subordinates to clients, it would be considered rude for restaurant-goers to say, for example, "Get me some water". Instead, restaurant-goers typically use indirect requests, such as "Could I have some water?". While this utterance may literally be a request for information, listeners effortlessly and instinctively infer the speaker's true intent, i.e., for the listener to bring them some water. Accordingly, so too must robots be able to infer the intended meanings behind their teammates' utterances according to their current context.

Research on indirect speech acts has a rich history in the philosophy of language literature dating back nearly half a century (Searle 1975), with computational approaches to indirect speech act understanding stretching back nearly that far (Perrault and Allen 1980). While early computational work focused on understanding indirect speech acts through inference-based plan-reasoning techniques, more recent work has either been cue-based, seeking to statistically estimate the illocutionary force of an utterance (e.g. whether an utterance is intended as a statement or request) without more deeply understanding the utterance's literal or intended meaning (Jurafsky 2004), or has been idiomatic in nature. This last category, the idiomatic approach, exploits the fact that many ISAs are so conventionalized that they are idiomatic (Wilske and Kruijff 2006), with ISA forms directly associated with inferred meanings, leading to significantly more computationally efficient handling of the vast majority of ISAs.

In previous work (Williams et al. 2014), we presented three capabilities necessary for a robust understanding of conventionalized ISAs through the idiomatic approach: (1) accounting for uncertainty in implication rules and their antecedent context, (2) fluidly adapting rules given new information, and (3) enabling better modeling of the beliefs of other agents. We also presented a Dempster-Shafer theoretic approach to ISA understanding (and generation, see Williams et al. 2015), which we argued increases the robustness of ISA inference by addressing the three capabilities mentioned above. This approach leverages a set of Dempster-Shafer Theoretic uncertain logical rules that map utterances to inferable intentions within specified contexts. These rules are annotated with Dempster-Shafer theoretic uncertainty intervals, denoting the amount of evidence for and against (and the degree of ignorance with respect to) each rule. But while we provided mechanisms for online adaptation of these rules when corrections are explicitly provided, we did not provide any mechanisms for initially learning these intervals from observation, requiring AI practition-

ers to initially handcraft rules' uncertainty intervals based on their own intuitions.

In this paper, we show how recent work on Dempster-Shafer Theoretic rule learning, originally developed and applied in the context of social and moral norm learning, can also be applied to learn Dempster-Shafer theoretic sociocultural politeness norms. After beginning with an overview of basic Dempster-Shafer theoretic concepts, we go on to provide a formal representation of ISA comprehension norms, a solution to solve the norm learning problem as applied to such norms, and the results of a series of experiments which provide the data necessary to use these norm learning algorithms. Finally, we present the results of our use of these algorithms to learn sociocultural linguistic politeness norms, and conclude with a discussion of possible directions for future work.

Dempster-Shafer Theory Background

Dempster-Shafer theory is a belief-theoretic uncertainty-processing framework (Shafer 1976), which has notions of belief and plausibility that are close to the inner and outer measures in probability theory (Fagin and Halpern 2013). Here, we introduce several basic notions in Dempster-Shafer Theory:

- **Frame of Discernment (FoD)**: A set of elementary events of interest is called a *Frame of Discernment* (FoD). A FoD Θ is a finite set of mutually exclusive events $\Theta = \{\theta_1, \cdots, \theta_n\}$. The power set of Θ is denoted by $2^\Theta = \{\mathcal{A} : \mathcal{A} \subseteq \Theta\}$.
- **Basic Belief Assignment (BBA)**: A *Basic Belief Assignment* is a mapping function $m_\Theta(\cdot) = 2^\Theta \rightarrow [0, 1]$ such that $\sum_{\mathcal{A}\subseteq\Theta} m_\Theta(\mathcal{A}) = 1$ and $m_\Theta(0) = 0$. The BBA measures the support assigned to the propositions $\mathcal{A} \subseteq \Theta$ only. The subsets of $\mathcal{A}$ with non-zero mass are referred to as *focal elements*, and comprise the core $\mathcal{F}_\Theta$. A *Body of Evidence* (BoE) is a triple $\varepsilon = \{\Theta, \mathcal{F}_\Theta, m_\Theta(\cdot)\}$.
- **Belief**, **Plausibility**, **& Uncertainty**: Given a BoE $\{\Theta, \mathcal{F}, m\}$, the *belief* of a set $\mathcal{A}(\mathcal{A} \subseteq \Theta)$ is $Bel(\mathcal{A}) = \sum_{\mathcal{B}\subseteq\mathcal{A}} m_\Theta(\mathcal{B})$. This belief function captures the total support that can be committed to $\mathcal{A}$ without also committing it to the complement $\mathcal{A}^c$ of $\mathcal{A}$. The *plausibility* of $\mathcal{A}$ is $Pl(\mathcal{A}) = 1 - Bel(\mathcal{A}^c)$. So, $Pl(\mathcal{A})$ corresponds to the total belief that does not contradict $\mathcal{A}$. The *uncertainty* interval of A is $[Bel(\mathcal{A}), Pl(\mathcal{A})]$.
- **Evidence Combination & Filtering**: The evidence from two sources having the BBAs $m_j(\cdot)$ and $m_k(\cdot)$ can be fused using various fusion strategies such as the Conditional Fusion Equation (CFE) (Premaratne et al. 2009) and evidence filtering strategies (Dewasurendra, Bauer, and Premaratne 2007).

Dempster-Shafer theory is often interpreted as extending traditional Bayesian frameworks (e.g., Pearl 2014) in several important ways. First, Dempster-Shafer theory makes weaker assumptions than Bayesian probability theory. For example, it does not require making any distributional assumptions. Second, it has the ability to directly express both *uncertainty*, which is represented in the mass function, and *ignorance*, which is represented in the amount of evidence not assigned to elementary events. Moreover, Dempster-Shafer Theory preserves this information during the evidence synthesis and belief combination process. Furthermore, Dempster-Shafer theory not only allows for the assignment of belief to a single elementary event within a FoD, but also to a subset of elementary events within that FoD, in order to expressing ambiguity between those elementary events. Finally, the Dempster-Shafer theory of evidence has uncertainty management and inference mechanisms in a way that is analogous to the human reasoning process at various levels of abstraction (Wu et al. 2002).

Norm Representation

In this section we present representations for *pragmatic rules*: sociocultural linguistic norms usable for both ISA understanding and generation. As a key component of our pragmatic rules and associated inference mechanisms, we adopt the representation of utterances proposed by Briggs and Scheutz (2011) (see also Briggs, Williams, and Scheutz 2017). Per Briggs and Scheutz (2011), an utterance u can be represented as:

$$u = UtteranceType(\alpha, \beta, X, M)$$

where $UtteranceType$ denotes the speech act type / illocutionary force of the utterance, α denotes the speaker, β denotes the hearer, X denotes the literal semantic meaning of the utterance, and M denotes a set of sentential modifiers (e.g., "now", "please") .

A *Norm* $\mathcal{N}$ is an expression of the form:

$$\mathcal{N} := u \wedge C \Rightarrow i$$

where *u* represents an utterance, *C* represents a possibly empty set of contextual conditions and *i* represents a possible intention that can be inferred from utterance *u* and contextual conditions *C*.

A sociocultural linguistic *Belief-Theoretic Norm* (cf. Sarathy, Scheutz, and Malle 2017) $\mathcal{N}$ is an expression of the form:

$$\mathcal{N} := u \wedge C \Rightarrow_{[\alpha,\beta]} i$$

where each norm is associated with a Dempster-Shafer theoretic uncertainty interval $[\alpha, \beta]$ $(0 \leq \alpha \leq \beta \leq 1)$.

Consider an example in which an agent X needs to determine Agent Y's intentions from their utterance "I could use some water" under two different contextual conditions: (1) X is working as a waiter in a restaurant where Y is a customer and (2) X and Y are working out together at a gym. We can represent this utterance and contextual conditions as follows:

$$u = Stmt(Y, X, could(use(Y, water))$$
$$C_1 = \{in(Y, restaurant), in(X, restaurant), customer(Y), waiter(X)\}$$
$$C_2 = \{in(Y, gym), in(X, gym), work\_out(Y), work\_out(X)\}$$

Then, we can leverage these intermediate representations in the formulation of *Belief-Theoretic Norms* in order to form a norm system as follows:

$\mathcal{N}_1 := u \wedge C_1 \Rightarrow$
$_{[0.2,0.3]}want(Y, believe(X, thirsty(Y)))$
$\mathcal{N}_2 := u \wedge C_1 \Rightarrow$
$_{[0.9,1.0]}want(Y, get\_for(Y, X, water))$
$\mathcal{N}_3 := u \wedge C_2 \Rightarrow$
$_{[0.75,0.95]}want(Y, believe(X, thirsty(Y)))$
$\mathcal{N}_4 := u \wedge C_2 \Rightarrow$
$_{[0.35,0.65]}want(Y, get\_for(Y, X, water))$

The norms in this example state that in the *restaurant* scenario or in the *exercise* scenario, Agent *X* can reasonably infer two intentions ("*Y wants X to believe that Y is thirsty*" and "*Y wants X to get Y some water*") from an utterance ("*I could use some water*") said by Agent *Y*. The lower and the upper bound of the interval associated with each norm indicates the level of support or evidence for that norm. Thus, in this example, $\mathcal{N}_1$ has a tight uncertainty interval with a low degree of belief, indicating the robot is certain that this norm does not apply; $\mathcal{N}_2$ with a tight uncertainty interval with a high degree of belief, indicating the robot is certain that this norm does apply; $\mathcal{N}_3$ also has a high degree of belief but a wider interval, indicating belief that the norm applies but with a higher degree of ignorance; and $\mathcal{N}_4$ has a wide uncertainty interval centered on zero, meaning that the robot has conflicting evidence for and against the norm holding, with a high degree of ignorance.

Norm Learning Problem

When using the proposed norms to infer intended meanings from ISAs, an agent must only consider the subset of norms in their norm system that apply in their current context. Using the previous norm system as an example, if Agent *X* knows that they are in a restaurant scenario, then *X* need only consider norms which are contextually applicable, such as $\mathcal{N}_1$ and $\mathcal{N}_2$. This subset, which Sarathy, Scheutz, and Malle (2017) define as a *norm frame* $\mathcal{N}_k^\Theta$, is a set of k norms, $k > 0$, in which every norm has the same utterance and the same set of contextual conditions. Thus, in the previous example, norms $\mathcal{N}_1$ and $\mathcal{N}_2$ would constitute one norm frame, while norms $\mathcal{N}_3$ and $\mathcal{N}_4$ would constitute another norm frame.

To learn these uncertainty intervals for sociocultural linguistic norms, we use the Dempster-Shafer theoretic norm learning algorithm presented by Sarathy et al. (2017). This algorithm takes as input a finite set of data instances, and updates the beliefs and plausibilities of candidate norms as it iterates over each data instance. Here, a data instance consists of a *norm frame* $\mathcal{N}_k^\Theta$, an evidence source s_i, a set of endorsements Φ_{si} provided by that source, and a mass assignment m_{si} corresponding to the amount of consideration or reliability placed on source s_i for this instance (Sarathy et al. 2017; Sarathy, Scheutz, and Malle 2017). While Sarathy et al. (2017) originally present this algorithm in the context of learning context-sensitive social and moral norms from human data, their approach is sufficiently general that we can easily apply it to our own norm learning problem. To do so, we begin by gathering data in a two-stage experimental process similar to that used by Sarathy et al. (2017).

Experiment 1: Generating Norms

In the experimental data collection phase, we followed the same experimental paradigm presented in Sarathy et al. (2017) and conducted two human subjects experiments. The first, the *generation* experiment, was used to identify candidate ISA understanding norms.

Method

We collected an initial set of candidate norms through an IRB-approved experiment using the psiTurk framework for Amazon Mechanical Turk (Gureckis et al. 2016). In this experiment, all participants began by providing informed consent to participate in the experiment. After completing a demographic questionnaire, participants then proceeded to the main experimental task, in which they were shown a sentence and asked to write down everything they could think of that the speaker might have meant by that sentence.

Because this work is focused on learning norms for interpreting conventionalized ISAs, each sentence shown to a participant followed a conventionalized ISA form, as shown in Table 1. These ISAs were chosen according to the taxonomies given by Briggs, Williams, and Scheutz (2017) and Searle (1975), and were phrased to evoke a context in which speakers regularly expect sociocultural politeness norms to be used when interacting with robots (Williams et al. 2018; Foster et al. 2012).

Experimental Results

163 U.S. participants were recruited from Amazon Mechanical Turk (76 female, 87 male; age range from 18 to 71, M=35.85, SD=15.14). After removing irrelevant responses, an average of 11.4 responses (min=7, max=17) were provided for each sentence. Ignoring responses with identical intentional interpretations, we gathered an average of 4.89 unique intentions (min=3, max=7) per sentence. Across all utterances, we collected 178 total responses, comprising 25 unique intentions.

Table 2 shows data collected for two different utterances ("Could I have some noodles?" and "I could use some noodles"), while Table 3 shows the five most frequent intentions across all collected data. As shown in those two tables, all collected intentions were converted from natural language to formal logical representations in order to systematically identify responses with identical semantic interpretations. For example, the semantic interpretation of the utterance "The speaker wants the hearer to get them some noodles" is represented as *want(S,get_for(H,S,noodles))*.

Experiment 2: Detecting Norms

Following Sarathy, Scheutz, and Malle (2017), our second experiment (the *detection* experiment) collected training data that could be used to learn uncertainty intervals for the norms identified in the previous experiment. Specifically, in this experiment, we collected human judgments as to whether different intentions were appropriate for different utterances in different scenarios. These scenarios were generated based on the sentences used in the first experiment and each scenario was a combination of an environmental

Question – Preparatory
Could you get some noodles?
Can you get some noodles?
Could I have some noodles?
Could I get some noodles?
Can I have some noodles?
Can I get some noodles?
Question – Sincerity
Were you going to get me the noodles?
Will you get me some noodles?
Would you mind getting me some noodles?
Statement – Sincerity
I need you to get me some noodles.
I would like you to get me some noodles.
I hope that you could get me some noodles.
I need some noodles.
I would like to order noodles.
I could use some noodles.
I hope I can get some noodles.
Statement – Preparatory
You can bring me some noodles.
You could bring me some noodles.
Suggestion – Preparatory
You should get me some noodles.

Table 1: Sentences used in the *generation* experiment. Yellow-highlighted sentences are agent-directed, while blue-highlighted sentences are patient-directed. From among these sentences, bolded sentences were selected to be used in the *detection* experiment.

context and a social context. All four experimental contexts are shown in Table 4.

Method

This experiment was conducted as a live, in-person laboratory study rather than on Mechanical Turk, as initial pilots of this experimental phase had difficulty obtaining high quality data from online participants. After providing informed consent, participants completed the main experimental task, followed by a demographic survey. In this experiment our demographic survey was taken after the main experimental task was complete because it contained questions which would have primed participants' interpretations of presented utterances (e.g., questions regarding experience in the restaurant industry).

The main experimental task took place over two rounds. In each of those two rounds, the participant was first introduced to an experimental context, in which they were either working as a waiter, or not, and in which they were either situated in a restaurant, or at a friend's house. We presented a different experimental context in each of the two rounds for every participant. Assignment of participants to pairs of experimental contexts was randomized. The decision to expose each participant to two of the four experimental contexts was made on the basis of time spent by participants during pilot testing.

In each round, after being given the experimental context, each participant was shown five sentences[1], each of which was followed by six candidate interpretations of that sentence. Participants were asked to select all interpretations from among those options that they believed to be interpretations of the presented sentence.

Intention Selection: For each selected sentence, the accompanying intentions presented to each participant included the two most frequently listed intentions for that sentence from the previous experiment, along with four other intentions randomly sampled from the distribution over all other intentions (with probability of selection corresponding to frequency of appearance in the previous experiment). For example, Table 5 shows the six possible intentions presented for the utterance "*Can you get some noodles?*", where the first two items were the two most frequently listed intentions for that utterance in the *generation* experiment and the remainder were sampled from the distribution over all other intentions.

Experimental Results

For this experiment, we recruited 37 participants from a college campus (14 female, 22 male, 1 NA, ages 18 to 36, M=21.03, SD=4.41). Participants came from 15 different departments, with the majority of participants coming from Mechanical Engineering (13 out of 37). 10 participants had previous experience in the restaurant industry. We collected 74 data points in total, with an average of 18.5 data points (min=17, max=20) collected in each context. In the next section, we describe how the data collected in this experiment were used to learn uncertainty intervals for the examined norms.

Learning Sociocultural Linguistic Norms from Human Data

To use the data collected in Experiment Two, we began by categorizing the data into subsets reflecting different norm frames. Specifically, four different types of norm frames were created based on the experimental contexts used in Experiment 2.

Type 1: Norm frames containing norms of the form:

$$u \wedge \emptyset \Rightarrow i$$

i.e., norms seeking to map utterance forms directly to intentions regardless of context. These rules' parameters can be learned from all data collected with respect to utterance u.

Type 2: Norm frames containing norms of the form

$$u \wedge \{[environmental\ context]\} \Rightarrow i$$

i.e., norms seeking to map utterance forms to intentions under particular environmental contexts (e.g., being in a

[1]Five sentences were selected from the total set of nineteen sentences to avoid overburdening participants. These sentences, shown in bold in Table 1, were selected to cover all five Speech Act taxonomic categories shown in that table. Because not all of these categories contained both agent-directed and patient-directed utterances, only agent-directed utterances were selected.

Could I have some noodles?		
Response	Logical Interpretation	Frequency
"The speaker wants the hearer to get them some noodles"	*want(S,get_for(H,S,noodles))*	5
"The speaker wants to order noodles from the hearer"	*want(S,order_from(S,H,noodles))*	4
"The speaker wants the hearer to share noodles with them"	*want(S,share_with(H,S,noodles))*	2
"The speaker wants the hearer to believe that the speaker is hungry"	*want(S,believe(H,hungry(S)))*	2
I could use some noodles		
Response	Logical Interpretation	Frequency
"The speaker wants the hearer to believe that the speaker is hungry"	*want(S,believe(H,hungry(S)))*	6
"The speaker wants the hearer to cook them some noodles"	*want(S,cook_for(H,S,noodles))*	4
"The speaker wants the hearer to get them some noodles"	*want(S,get_for(H,S,noodles))*	2

Table 2: A subset of results from the *generation* experiment, collected for two different utterances.

Response	Logical Interpretation	Frequency
"The speaker wants the hearer to get them some noodles"	*want(S,get_for(H,S,noodles))*	59
"The speaker wants the hearer to buy them some noodles"	*want(S,buy_for(H,S,noodles))*	24
"The speaker wants to order noodles from the hearer"	*want(S,order_from(S,H,noodles))*	24
"The speaker wants the hearer to believe that the speaker is hungry"	*want(S,believe(H,hungry(S)))*	17
"The speaker wants the hearer to cook them some noodles"	*want(S,cook_for(H,S,noodles))*	17

Table 3: The five most frequent intentions from overall data in the *generation* experiment.

	Restaurant	Friend's House
Waiter	Scenario 1	Scenario 2
Not Waiter	Scenario 3	Scenario 4

Table 4: Scenarios in the *detection* experiment. Two contextual conditions are *in a restaurant/in a friend's house* (environmental contexts) and *work as a waiter/do not work as a waiter* (social contexts).

restaurant). These rules' parameters can be learned from all data collected with respect to utterance u under that particular environmental context.

Type 3: Norm frames containing norms of the form

$$u \wedge \{[social\ context]\} \Rightarrow i$$

i.e., norms seeking to map utterance forms to intentions under particular social contexts (e.g., the listener is a waiter). These rules' parameters can be learned from all data collected with respect to utterance u under that particular social context.

Type 4: Norm frames containing norms of the form

$$u \wedge \{[environmental\ context] \wedge [social\ context]\} \Rightarrow i$$

i.e., norms seeking to map utterance forms to intentions under particular combinations of environmental and social context. These rules' parameters can be learned from all data collected with respect to utterance u under that particular combination of environmental and social context.

Under the above criteria, we gathered 5 Type 1 subsets (corresponding to the five utterances) with 74 data instances for each norm frame, 10 Type 2 subsets (corresponding to the five utterances and two environmental contexts) with 37 data instances for each norm frame, 10 Type 3 subsets (corresponding to the five utterances and two social contexts) with an average of 37 (min=35, max=39) data instances for each norm frame, and 20 Type 4 subsets (corresponding to the five utterances, two environmental contexts, and two social contexts) with an average of 18.5 (min=17, max=20) data instances for each norm frame. Note that as we progress through these types of norm frames, we attempt to learn increasingly context-specific norms from increasingly limited datasets. For each of these norm frames, we provided the data commensurate with that norm frame to Sarathy, Scheutz, and Malle (2017)'s rule learning algorithm in order to learn uncertainty intervals for each norm in that frame.

Table 6 shows, as an example, one of the Type 3 norm frames generated through this process, using the utterance "*Can you get some noodles?*" in the social context that the hearer works as a waiter, with no restrictions on environmental context. The first column lists norms candidates applicable within this norm frame. The second column shows consensus from participants, i.e., what percentage of participants selected the interpretation associated with each norm under the specified utterance and context. The last column shows the Dempster-Shafer theoretic uncertainty intervals generated after feeding all applicable training data to the rule learning algorithm.

Norm Selection

After learning uncertainty intervals for each candidate norm, we must next select a subset of "justifiable" norms that should be included in the final norm system. This can be achieved by only accepting norms that reflect a sufficient level of confidence in the consequent given the antecedent. In this paper we will examine the effects of selecting only the subset of norms for which the center of the norm's learned

Candidate Intention	Logical Interpretation
The speaker wants the hearer to get them some noodles.	Int_1 = *want(S,get_for(H,S,noodles))*
The speaker wants the hearer to buy them some noodles.	Int_2 = *want(S,buy_for(H,S,noodles))*
The speaker wants to order noodles from the hearer.	Int_3 = *want(S,order_from(S,H,noodles))*
The speaker wants the hearer to believe that the speaker is hungry.	Int_4 = *want(S,believe(H,hungry(S)))*
The speaker is asking the hearer for permission whether they can have noodles.	Int_5 = *ask(S,permission(H,have(X,noodles)))*
The speaker wants the hearer to cook them some noodles.	Int_6 = *want(S,cook_for(H,S,noodles))*

Table 5: Candidate intentions presented for the utterance "*Can you get some noodles?*" in the *detection* experiment, along with the logical representations of those candidate intentions.

Norm	Consensus	Interval
$u \wedge C \Rightarrow Int_1$	0.838	[0.572,0.840]
$u \wedge C \Rightarrow Int_2$	0.432	[0.343,0.610]
$u \wedge C \Rightarrow Int_3$	0.622	[0.453,0.721]
$u \wedge C \Rightarrow Int_4$	0.432	[0.308,0.576]
$u \wedge C \Rightarrow Int_5$	0.27	[0.194,0.461]
$u \wedge C \Rightarrow Int_6$	0.27	[0.024,0.291]

Table 6: Result of a norm frame using utterance $u = QuestionYN(S, H, can(H, get(X) \wedge noodles(X)), \{\})$, i.e., "*Can you get some noodles?*" with social context $C = waiter(H)$, i.e., "the hearer work as a waiter" and inferred intentions as listed in Table 5.

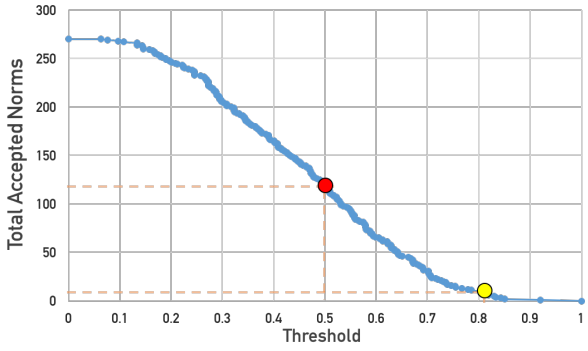

Figure 1: Results from the norm learning algorithm, shows how the total of accepted norms decreases as the threshold τ increases. The red dot shows that with threshold 0.5 (selecting norms whose uncertainty intervals are centered at a point greater than 0.5) 115 total norms are selected for adoption; the yellow dot shows that when a threshold of 0.81 is used, only 6 are selected for adoption.

uncertainty interval (i.e., $\frac{\alpha+\beta}{2}$) is above some threshold τ. Figure 1 shows how the total of accepted norms declines as this threshold is reduced from 115 with threshold $\tau = 0.5$ to 6 with threshold $\tau = 0.81$. If a lower value of τ is chosen, a greater number of norms will be learned, but the agent may need to generate a greater number of clarification requests in the future; if a higher value is chosen, fewer norms will be learned but the agent may need to generate fewer clarification requests in the future[2].

Figure 2 shows how Sarathy, Scheutz, and Malle (2017)'s norm learning algorithm converges to different uncertainty intervals for each of the norms selected with a threshold of $\tau = 0.81$. Notice here that the amount of ignorance (i.e., $\beta - \alpha$) converged to for each norm is directly determined by the number of terms in that norm's antecedent, as this directly determines the number of datapoints of evidence provided to the algorithm.

Discussion and Conclusion

In this paper, we presented the first approach to automatically learning the confidence intervals in Dempster-Shafer theoretic work on language understanding and generation, by demonstrating how to apply recent work on Dempster-Shafer Theoretic rule learning (Sarathy, Scheutz, and Malle 2017) to learning appropriate uncertainty intervals for robots' representations of sociocultural politeness norms. Unlike Sarathys approach, in which norm antecedents only contain a single context label, we show how this algorithm can be used with rules whose left-hand sides contain logically specified context descriptions of varying levels of specificity. We also demonstrated how sets of candidate norms can be selected based on the characteristics of their learned confidence intervals and how this selection process can be used to select norms that have an appropriate level of contextual specificity.

While we have demonstrated how sensible uncertainty intervals for sociocultural linguistic norms can be learned offline from human data, a number of challenges remain for future work. Two key challenges for future work are how to identify and generate candidate ISA understanding norms, and how to adapt those norms to facilitate lifelong learning. In this work, we selected a relatively narrow domain along with utterances that we explicitly expected to be generated due to sociocultural linguistic norms relevant to that domain. In the future it will be important to extend the methods presented in this paper to work online for long timescales over the utterances that occur naturally in human-robot dialogue, and to develop mechanisms for proposing candidate norms based on salient aspects of the robot's context. In previous work, Williams et al. (2014; 2015) discussed how Dempster-Shafer theoretic norm representations could theoretically be updated online using the Conditional Update Equation (Premaratne et al. 2009), but to the best of our knowledge there

[2]Note that other threshold options are possible: selecting a threshold based on α will only select rules that have sufficient evidence in favor of them; selecting a threshold based on β will only select rules that lack sufficient evidence against them.

$\mathcal{N}_1$	$QuestionYN(S, H, will(H, get(X) \wedge noodles(X))) \Rightarrow_{[0.88,0.96]} want(S, get\_for(H, noodles)$
$\mathcal{N}_2$	$QuestionYN(S, H, will(H, get(X) \wedge noodles(X))) \wedge \neg waiter(H) \Rightarrow_{[0.73,0.97]} want(S, get\_for(H, noodles)$
$\mathcal{N}_3$	$Stmt(S, H, need(H, get(S, noodles))) \wedge waiter(H) \Rightarrow_{[0.70,0.99]} want(S, get\_for(H, noodles)$
$\mathcal{N}_4$	$QuestionYN(S, H, will(H, get(X) \wedge noodles(X))) \wedge \neg in(H, restaurant) \Rightarrow_{[0.70,0.97]} want(S, get\_for(H, noodles)$
$\mathcal{N}_5$	$Stmt(S, H, should(H, get(S, noodles))) \Rightarrow_{[0.79,0.87]} want(S, get\_for(H, noodles)$
$\mathcal{N}_6$	$Stmt(S, H, need(H, get(S, noodles))) \wedge in(H, restaurant) \Rightarrow_{[0.68,0.95]} want(S, get\_for(H, noodles)$

Table 7: Learned Norms. This table shows the six norms, $\mathcal{N}_{1...6}$ selected by our approach when a threshold of $\tau = 0.81$ was used. The convergence trajectories for each norm is shown below in Fig. 2.

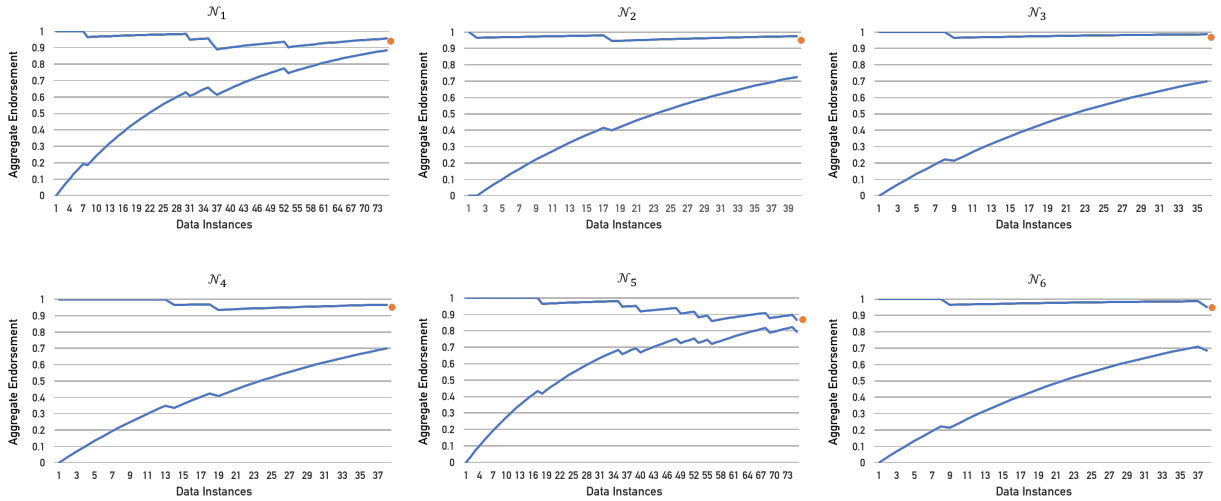

Figure 2: Results from the norm learning algorithm, showing convergence to different uncertainty intervals for each of the norms selected with a threshold of $\tau = 0.81$. The dots represent the mean norm endorsements by experimental participants.

has been no approach to date that actually triggers the use of this equation based on robot observations, nor from statements, clarifications, or corrections made by human interlocutors.

Future work must also demonstrate the efficacy of learned norms on physical robotic platforms in interactions with real users. In future work, we plan to encode our learned norms within the pragmatic norm base (cf. Williams et al. 2015) used by the Distributed Integrated Affect, Recognition and Cognition (DIARC) architecture (Schermerhorn et al. 2006; Scheutz et al. 2013; 2019) and assess the fluidity and successfulness of robots interacting with users under norm systems selected with different thresholds.

Finally, we must consider how this norm-based approach fits into the broader context of ISA understanding and generation. The sociocultural linguistic norms learned in this paper are learned specifically for use within an idiomatic approach to ISA processing. In future work, it will be critical to integrate this approach with a plan-reasoning system capable of handling unconventionalized indirect speech acts as well, such as that presented by Hinkelman and Allen (1989) (cp. Briggs and Scheutz 2013).

Acknowledgments

This work was supported in part by NSF grants IIS-1849348 and IIS-1909847.

References

Briggs, G., and Scheutz, M. 2011. Facilitating mental modeling in collaborative human-robot interaction through adverbial cues. In *Proceedings of the SIGdial Meeting on Discourse and Dialogue (SIGDIAL) 2011 Conference*, 239–247.

Briggs, G. M., and Scheutz, M. 2013. A hybrid architectural approach to understanding and appropriately generating indirect speech acts. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Briggs, G.; Williams, T.; and Scheutz, M. 2017. Enabling robots to understand indirect speech acts in task-based interactions. *Journal of Human-Robot Interaction (JHRI)*.

Dewasurendra, D. A.; Bauer, P. H.; and Premaratne, K. 2007. Evidence filtering. *IEEE Transactions on Signal Processing* 55(12):5796–5805.

Fagin, R., and Halpern, J. Y. 2013. A new approach to updating beliefs. *arXiv preprint arXiv:1304.1119*.

Foster, M. E.; Gaschler, A.; Giuliani, M.; Isard, A.; Pateraki, M.; and Petrick, R. 2012. Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, 3–10. ACM.

Gureckis, T.; Martin, J.; McDonnell, J.; et al. 2016. psiturk: An open-source framework for conducting replicable

behavioral experiments online. *Behavior Research Methods* 48(3):829–842.

Hinkelman, E. A., and Allen, J. F. 1989. Two constraints on speech act ambiguity. In *Proceedings of ACL*.

Jurafsky, D. 2004. Pragmatics and computational linguistics. *The handbook of pragmatics* 578.

Pearl, J. 2014. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Perrault, C. R., and Allen, J. F. 1980. A plan-based analysis of indirect speech acts. *Computational Linguistics* 6(3-4):167–182.

Premaratne, K.; Murthi, M. N.; Zhang, J.; Scheutz, M.; and Bauer, P. H. 2009. A Dempster-Shafer theoretic conditional approach to evidence updating for fusion of hard and soft data. In *2009 12th International Conference on Information Fusion*, 2122–2129. IEEE.

Sarathy, V.; Scheutz, M.; Kenett, Y. N.; Allaham, M.; Austerweil, J. L.; and Malle, B. F. 2017. Mental representations and computational modeling of context-specific human norm systems. In *CogSci*, volume 1, 1–1.

Sarathy, V.; Scheutz, M.; and Malle, B. F. 2017. Learning behavioral norms in uncertain and changing contexts. In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 000301–000306. IEEE.

Schermerhorn, P. W.; Kramer, J. F.; Middendorff, C.; and Scheutz, M. 2006. DIARC: A testbed for natural human-robot interaction. In *AAAI*, volume 6, 1972–1973.

Scheutz, M.; Briggs, G.; Cantrell, R.; Krause, E.; Williams, T.; and Veale, R. 2013. Novel mechanisms for natural human-robot interactions in the DIARC architecture. In *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Scheutz, M.; Williams, T.; Krause, E.; Oosterveld, B.; Sarathy, V.; and Frasca, T. 2019. An overview of the distributed integrated cognition affect and reflection DIARC architecture. In *Cognitive Architectures*. Springer. 165–193.

Searle, J. R. 1975. Indirect speech acts. *Syntax and Semantics* 3:59–82.

Shafer, G. 1976. *A mathematical theory of evidence*, volume 42. Princeton university press.

Williams, T.; Núñez, R. C.; Briggs, G.; Scheutz, M.; Premaratne, K.; and Murthi, M. N. 2014. A dempster-shafer theoretic approach to understanding indirect speech acts. In *Ibero-American Conference on Artificial Intelligence*, 141–153. Springer.

Williams, T.; Briggs, G.; Oosterveld, B.; and Scheutz, M. 2015. Going beyond command-based instructions: Extending robotic natural language interaction capabilities. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Williams, T.; Thames, D.; Novakoff, J.; and Scheutz, M. 2018. "Thank you for sharing that interesting fact!": Effects of capability and context on indirect speech act use in task-based human-robot dialogue. In *Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

Wilske, S., and Kruijff, G.-J. 2006. Service robots dealing with indirect speech acts. In *Proc. of the Int'l Conference on Intelligent Robots and Systems*.

Wu, H.; Siegel, M.; Stiefelhagen, R.; and Yang, J. 2002. Sensor fusion using dempster-shafer theory [for context-aware hci]. In *IMTC/2002. Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 00CH37276)*, volume 1, 7–12. IEEE.