“I think it might help if we multiply, and not add” : Detecting Indirectness in Conversation

Pranav Goel, Yoichi Matsuyama, Michael Madaio and Justine Cassell

Abstract The strategic use of indirect language is crucial in business negotiations, doctor-patient interactions, instructional discourse and multiple other contexts. Being indirect allows interlocutors to diminish the potential threats to their interlocutors' desired self image - or, face threat - that may arise by being overtly direct. Handling indirectness well is important for spoken dialogue systems, as being either too indirect or too direct at the wrong time could be harmful to the agent-user relationship. We take a step towards handling users' indirection by exploring the task of automatically detecting *indirectness* in conversations, exploring different supervised machine learning approaches, and ultimately achieving a 62% F1 score on our dataset. Deep neural network based approaches perform significantly better than their non-neural counterparts, which may indicate the nuanced and complex nature of indirectness. To our knowledge we are the first to use a multi-modal approach to detecting indirect language: we rely on both verbal and nonverbal features of the interaction. Accurate automated detection of indirectness may help conversational agents better understand their users' intents, gauge the current relationship with the user in order to appropriately plan a response, and inform the strategic use of indirectness to manage the task goals and social goals of the interaction.

1 Introduction

Indirect delivery, or indirectness, is the linguistic phenomena where the speaker intentionally does not communicate their intention straightforwardly. This is done

Pranav Goel
Department of Computer Science & Engineering, Indian Institute of Technology (BHU) Varanasi
e-mail: pranav.goel.cse14@iitbhu.ac.in

Yoichi Matsuyama, Michael Madaio and Justine Cassell
School of Computer Science, Carnegie Mellon University
e-mail: {yoichim,mmadaio,justine}@cs.cmu.edu

either by distancing themselves from their intended meaning or the proposition they are communicating, by introducing vagueness, reducing certainty or intensity, or making their statements appear more subjective, among others [41, 40, 11, 36, 27]. It is a conversational strategy that is seen as intrinsic to conversational discourse [29]. In a variety of linguistic contexts, such as negotiation, counseling, health care [48], and education [38, 27], it is often important to speak indirectly to communicate most effectively. The stakes in such conversations are often quite high, ranging from profitable business deals to life-saving medical advice. In many cases, delivering information (e.g. a dire medical diagnosis) in a direct manner, while perhaps successful in communicating information in the short-term, can have serious consequences for the long-term relationship and even the success of the interaction. The strategic use of *indirectness* is thus critical to softening the blow of such direct delivery, resulting in more effective communication [4]. In business negotiation, for example, being indirect is used to present tentative views, weaken one's commitment to a particular bid, and contribute to trust between negotiators [49].

Apart from using indirectness strategically, interlocutors are expected to *detect* it in each other's utterances. Consider bosses mitigating orders in office meetings to sound less threatening and more friendly [7]. The employee is expected to parse the request despite its indirectness and understand that it is actually imperative to get the requested work done. Computer-mediated discourse and the use of virtual assistants is increasing in many domains, and their use in the domains above makes the issue of detecting indirectness particularly important [37, 24, 18]. Knowing when a user is being indirect can help such systems better understand, respond to, and build the user-agent relationship to help them be more effective in the long run.

The prevalence and importance of indirectness in educational interactions [35] coupled with the rise of tutoring dialogue systems such as "teachable agents" [18] make education a productive domain for us to target in this work. Students use such indirectness when proposing answers to their teachers [38], and students peer tutoring one another [32] use indirectness to either communicate uncertainty or to reduce the threat to their partner's self image and self-esteem that might result from overly direct feedback [35, 27]. Without such indirectness, peer tutors' requests and feedback may threaten what Brown and Levinson (following Goffman) call one's "positive face", or desire to be seen in a positive light [4]. [27] found that peer tutors who had greater self-efficacy (i.e. confidence in their tutoring ability) used more indirectness, suggesting that this plays a strategic face-saving role, and thus a relationship-building role. Therefore, intelligent tutoring systems, whether they are playing the role of a tutor or of a student (in the case of a teachable agent), particularly those that attempt to build a social motivational relationship with students, would benefit from detecting indirectness used by the students. It is also notable that [27] found that tutors with a stronger relationship, or "rapport" with their partners were more direct. In a spoken dialogue system, if a user is being more direct, having the agent be continuously polite (as in [19]) may in fact be perceived as distancing and may harm the rapport between the agent and user [35] (note that being polite is one function of indirectness, but politeness and indirectness are different phenomena as explained in Section 2). Automatically detecting indirectness in user

utterances can thus help all kinds of spoken dialogue systems better estimate the state of the social relationship - an aid in designing an appropriate response - as well as more effectively communicating information.

To detect indirectness in conversation, we use various machine learning approaches such as Support Vector Machines and Logistic Regression (for which we try various feature combinations). This is motivated in part by studies carried out by [36], where they observed and gave empirical proof for the lack of effectiveness of a simple keyword search based approach for uncertainty or hedge detection. We also use several neural network based methods, which have been shown to give improved performance over non neural baselines for many NLP applications including detection of uncertainty in texts [1] and politeness classification in requests [2], both of which are related to, though distinct from, indirectness (section 2). To the best of our knowledge, this is the first work to leverage machine learning methods for automated detection of indirectness in dialogue. We take the first step towards multimodal detection of indirect language in conversations by leveraging both the verbal (text from dialogue transcript) and visual (features extracted from video recordings) modalities. Our approach and results can help lead to automatic indirectness detection used in a variety of spoken dialogue systems, from tutoring dialogue systems, to socially-aware conversational agents.

2 Related Work

Prior work on automated detection of indirectness has focused only on a specific function of indirectness (such as politeness [3]), or targeted some specific manifestation of the phenomena (such as uncertainty [31, 12]) using only a single modality (i.e. text) [11, 36, 8]. The crucial difference is that in our work, we attempt to detect indirectness, and not just a specific function or manifestation, and try to do so by leveraging multiple modalities together. For example, [8] explored a politeness classifier based on syntactic and lexical features, incorporating various components of politeness theory. [2] then used neural networks for the same task on the same corpus. Their annotation for politeness includes being indirect as *one of the ways* of being polite in requests. However, indirectness is not always interpreted as politeness and can even be associated with lack of politeness [3] (consider example 1 in Table 1). Indirectness has many more functions in addition to marking politeness [31, 7], while one can be polite in ways other than being indirect [8] (example 3 in Table 1). Indirectness is often produced through the use of hedges, which are "single- or multi-word expressions used to indicate uncertainty about the propositional content of an utterance or to diminish its impact" [31, 12]. Thus, uncertainty is just one of the ways in which indirectness can manifest in conversations (see examples 1 and 2 in Table 1). A statement may also be uncertain (due to lack of exact percentages or numeric data when trying to quantify something) without being indirect (consider example 4 in Table 1).

Many studies in NLP have explored the detection of such hedges, focusing only on the uncertainty aspect [11, 36], especially for text. The ConLL 2010 shared task on hedge or uncertainty detection [11] facilitated automated separation of 'uncertain' and 'factual' statements by providing two annotated datasets - a BioScope corpus (abstracts and articles from biomedical literature) and a Wikipedia corpus. Recently, [1] carried out deep neural network based experiments on the ConLL 2010 shared task datasets for uncertainty detection to explore different kinds of attention mechanisms using the task setting. However, text-based corpora don't allow for use of the rich data communicated by nonverbal behavior.

Even for spoken dialogue settings, past work has again focused on 'uncertainty' detection and not the broader phenomena of being indirect. [26] used prosody to automatically detect student 'certainness' in spoken tutorial dialogue. [9] also investigated automatic detection of uncertainty using predefined prosodic markers. If the targeted prosodic markers could not be identified for a certain utterance, they fell back on a defined list of lexical markers to classify an utterance as certain or uncertain. The phenomena of indirectness we study in our corpus relates more to the general face threat mitigation needs in dialogue rather than simply a way of introducing uncertainty [27]. Prior work on spoken dialogue and text based corpora primarily relied on just one modality, or only using one modality at a time when classifying an instance. While verbal (text) and vocal (speech) modalities have been explored, no past work has leveraged the 'visual' modality to the best of our knowledge. The use of nonverbal behaviors (including visual features like eye contact, smiling, and more) has been motivated by [9] for uncertainty detection (based on the experiments carried out by [23]), and by [44] as crucial for face threat mitigation.

#	Example	Indirect	Polite	Uncertain
1	can you please ***just*** stay with me and not doodle	✓	✗	✗
2	er A equals twenty-four ***sorry***	✓	✓	✗
3	***Nice work*** so far on your rewrite.	✗	✓	✗
4	The club enjoyed ***most of its success*** in its early years.	✗	✗	✓

Table 1 Examples showing how politeness, uncertainty and indirectness are different phenomena. The first two examples are from our reciprocal peer tutoring corpus (see Section 3), example 3 from the corpus of requests annotated for politeness by [8] and example 4 from the Wikipedia corpus annotated for uncertainty detection by [11].

3 Corpus Collection and Annotation

Our dialogue corpus was collected from 12 American-English speaking pairs (or dyads) of teenagers (mean age = 13.5) tutoring each other in basic linear algebra. They worked together for 5 weekly hour-long sessions for a total of about 60 hours.

Code	Definition	Example	Distribution
Apology	Apologies used to soften direct speech acts	Sorry, its negative 2.	7.7%
Qualifiers	Qualifying words for reducing intensity or certainty	You just add 5 to both sides.	66.1%
Extenders	Indicating uncertainty by referring to vague categories	You have to multiply and stuff.	3.6%
Subjectivizer	Making an utterance seem more subjective to reduce intensity	I think you divide by 3 here.	22.6%

Table 2 Annotation of codes under the 'indirect' label. Distribution = % of all indirect utterances.

Each session included some social interaction as well as one of the members of the dyad tutoring the other (the roles are reversed later in the session). Indirectness was annotated only for the 'tutoring' periods. Audio and video data were recorded, transcribed, and segmented for clause-level dialogue annotation of various conversational strategies including *indirectness* or indirect delivery. The corpus was coded for four types of indirectness - apologizing, hedging language (e.g. use of qualifiers), the use of vague category extenders, and "subjectivizing" [12, 31]. These are detailed in Table 2. For all the four codes, the Krippendorff's alpha for five trained annotators was at least 0.7. Once the annotators reached sufficient inter-rater reliability, the corpus was divided amongst the annotators, each labeling one fifth of the corpus. An utterance was classified as indirect or not based on its inclusion in any of these four categories. After some data cleaning and simple pre-processing steps (not detailed here for brevity), we have a total of 23437 utterances, with 1113 out of them labeled as 'indirect'.

- Eye Gaze - Three types of gaze were annotated - Gaze at Partner (gP), Gaze at the worksheet (gW), and Gaze elsewhere (gE).
- Smile - A smile is defined by the elongation of the participants lips and rising of their cheeks. Smiles were annotated from the beginning of the rise to the end of the decay (as per the parameters explained in [17]). Laughter (including smiling) has shown to be an effective method of face threat mitigation [45], and therefore might be used in conjunction with indirect language.
- Head Nod - Temporal intervals of head nod were encoded (beginning of the head moving up and down until the moment the head came to rest).

Inter-rater reliability for visual behavior was 0.89 for eye gaze, 0.75 for smile count (how many smiles occur), 0.64 for smile duration and 0.99 for head nod. Further details of extraction and ground truth definitions for each behavior can be found in [50], who found these behaviors were important for automatic detection of social norm violation, self-disclosure, praise and reference to shared experience in conversations. In particular, they found gaze behaviors, head nods, and smiling helpful. In addition, [20] showed that head tilt was one of the strongest nonverbal cue to interpersonal intimacy.

4 Approaches

Our (supervised) ML methods include non neural network based approaches relying on various sets of features and different neural network based architectures (due to their use in related tasks like politeness and uncertainty detection, as mentioned in section 1). We mention the variations, but focus only on the best performing models.

4.1 Non neural network based

These methods involve feature representation and training the learning algorithm.
Feature Representation: We tried various ways to represent the utterances. Table 3 summarizes the features we considered. We briefly discuss these features below -

- Unigram or bag-of-words: We set a rare threshold of 10 for our experiments, which means that a word (or the target n-gram) will be considered only if it occurs at least 10 times in the training set.
- Pair-based features: To capture some context beyond just the words, we use a feature representation consisting of bigrams, Part-of-Speech (POS) bigrams, and word-POS pairs (also used by [50] for conversational strategy classification). The rare threshold is again set to 10.
- Pre-trained word vectors: Word2Vec [30] and GloVe [34] are useful ways to represent text in a vector space of manageable dimensionality. This methodology has been successfully applied to many NLP tasks [10, 30]. We tried various available pre-trained models like *Twitter word2vec* [14] trained on 400 million Twitter tweets, GloVe representations [34] trained on Wikipedia articles (called *GloVe wiki*) and web content crawled via Common Crawl (called *GloVe Common Crawl*), and word vectors trained on Wikipedia articles using Word2Vec by [1], referred to as *Wikipedia Word2Vec* by us.
- Word2Vec trained on our dataset: We learn word vector representations on our own training data. We refer to this model as *RPT Word2Vec* (for Reciprocal Peer Tutoring). We tune the various training parameters on the validation dataset (see Training Detail) resulting in window size = 9, dimensionality = 300 and training algorithm = continuous bag-of-words.
  Note that for word embeddings, we get the representation of each word of a sentence. To get the representation for the overall sentence (to apply non neural ML algorithms), we take the *mean* of the individual word embeddings.
- Visual features: Visual behaviors annotated for our corpus were explained in Section 3. Three types of eye gaze, smile and head nod were annotated for both the tutor and tutee at each turn, giving a set of 10 visual features (Table 3).

Training Detail: Our corpus contains 60 sessions of peer tutoring interaction. Out of these, we take 48 sessions as the training dataset, 6 as validation and 6 as test set. This is repeated 5 times to get five train-validation-test splits. We use the same splits across every experiment. The splits were done on the basis of dialogue sessions

Extracted from text		**Extracted from video**
Bag-of-words/n-gram	Word embeddings	
Unigram (∼900)	Twitter Word2Vec (400)	Visual (10)
Pair-based [Bigrams, POS bigrams, Word-POS pairs] (∼3700)	GloVe wiki (300)	
	GloVe Common Crawl (300)	
	Wikipedia Word2Vec (400)	
	RPT Word2Vec (300)	

Table 3 Summary of the feature sets (dimensionality) used to represent utterances

across all speaker dyads. The validation set is used to decide the best performing approaches, tune hyperparameters and to choose the training settings for Word2Vec. For each feature representation, we tried the following supervised machine learning algorithms - Logistic Regression, Naive Bayes, Random Forest and Support Vector Machine (SVM), with Logistic Reg. and SVM performing best on the validation set.

4.2 Neural Network based

We applied various neural architectures to see if they could perform better than the non-neural models for indirectness detection. Since this task has not been explored directly in the past (section 2), we tried many different architectures that have worked well in past classification-based NLP work. These include fully connected or feedforward neural networks, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) (including variants like Long Short-Term Memory (LSTM) [16] network and Gated Recurrent Units [5]) which have all been applied successfully to various NLP tasks [6, 15, 25] including politeness classification and hedge detection [2, 1]. CNNs and RNNs expect sequential input, and hence we concatenate the word vector representations to form the sentence representation. Combining LSTMs and CNNs in sequence, and having one or more fully connected layers after convolutional/recurrent layers (Figure 1) has also proven to be effective in NLP [51, 25]. For pooling the sequential output of CNN/LSTM, we tried taking the maximum, mean, or only the final vector in the sequence.

In our experiments, we incorporated visual features by concatenating the feature vector (of dimensionality 10, see Table 3) with the input to the first feedforward or fully connected layer (which could also be the output layer) of the deep neural network (Figure 1) for all the various architectures we tried. This method is inspired from other similar ways to incorporate external features in neural architectures [33]. Recently, the attention mechanism has been successfully applied to augment CNNs and LSTMs [46, 47]. Certain portions of a sequence are more predictive of the output than others, and the selective mechanism of attention allows the network to focus on the most relevant parts of an input or hidden layer, which is useful for long input sequences. We tried all these various networks and compared their performances on the validation set. The two best performing architectures for indirectness detection are discussed below.

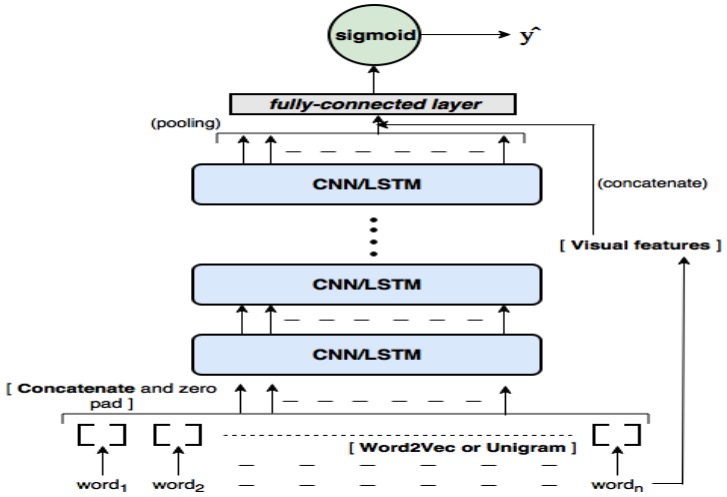

Fig. 1 A general representation of the different neural architectures (and combinations) tried

4.2.1 Stacked LSTMs

Stacking multiple LSTM layers one after the other has been an effective method for various NLP tasks such as dialogue act classification [21] (which is close to our task since we are classifying a conversational strategy in dialogue). Some of the architectural decisions include the number of layers to be stacked, having a fully connected layer after the stacked LSTMs or not, the activation for output layer, etc.

We experimented with two common choices for the final output layer - softmax, which gives a probability distribution over the number of classes or a single neuron and sigmoid activation, which gives a real valued score between 0 and 1 ($\hat{y}$ in Figure 1). Using a sigmoid function requires choosing a threshold value, such that an utterance assigned score above this value gets labeled '1' (presence of indirectness). We found that sigmoid worked better for our task based on validation set results, and tune the threshold on validation set. The sequential output from the final LSTM was pooled using last pooling. This vector can then be concatenated with the vector representing the visual features, before going to either some fully connected layers (Figure 1) or to the output layer itself. We applied Dropout [42] at each LSTM layer, tuning the dropout rate on validation set. After fine-tuning these hyperparameters, we get the best performing setting reported in Table 4. The network parameters for the neural model were learned by minimizing the binary cross-entropy loss [39] between the actual and predicted labels. We optimized this function by back-propagating through layers via Mini-Batch Gradient Descent using a batch size of 512, 25 training epochs and Adam optimization algorithm [22] with the parameters set as $\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999$ *and* $\varepsilon = 10^{-9}$.

	Initial Embedding Layer	#Stacked LSTM layers (dimensionality)	Dropout Rate	Sigmoid Threshold
RPT dataset	Twitter Word2Vec	4 (400, 300, 200, 100)	0.5	0.3
BioScope	Wikipedia Word2Vec	3 (400, 200, 100)	0.2	0.4
Wikipedia	GloVe wiki	4 (400, 300, 200, 100)	0.5	0.3

Table 4 Hyperparameter settings for the Stacked LSTMs approach which resulted in the best validation set performance

4.2.2 Attention-based CNN

[1] applied attention mechanism in different ways to the task of uncertainty detection on the ConLL 2010 shared tasks datasets [11]. They outperformed the best shared task system on the Wikipedia dataset while matching the state-of-the-art on the Biomedical dataset (Table 6). We tried using their methodology on our peer-tutoring dataset for indirectness detection. As per [1], the attention layer a for input x is given as: $\alpha_i = \frac{exp(f(x_i))}{\sum_j exp(f(x_j))}; a_i = \alpha_i \cdot x_i$ where f is a scoring function, the α_i are attention weights and each input x_i gets re-weighted (selectively focused upon) by its corresponding attention weight α_i The most basic definition for f is the linear scoring function (on the input x): $f(x_i) = W^T x_i$. W are parameters learned in training. We applied attention on the input sequence itself (Att_Inp CNN) and on the hidden layer of the convolution (Att_Conv CNN) (for details, refer [1]). We also tried using attention on LSTM which was not as effective as using CNN.

5 Results and Discussion

We use the F1 score as the evaluation metric to measure the performance of our various models. Since our dataset is unbalanced (or skewed towards the '0' class which means absence of indirectness), accuracy would not be a good choice. F1 score was also used in the ConLL 2010 shared task. The best performing systems on ConLL 2010 shared task on uncertainty detection [11] used essentially SVM on bag-of-words based features, for both the Wikipedia and BioScope datasets [13, 43].

	Logistic Reg.	SVM
Unigram	57.71	59.1
Unigram+Visual	57.74	**59.3**
Pair-based	57.09	58.28
Pair-based+Visual	55.89	58.41
Twitter Word2Vec	44.83	53.86
GloVe Wiki	37.91	45.25
GloVe Common Crawl	38.94	45.06
Wikipedia Word2Vec	44.56	49.54
RPT word2vec	44.95	39.36

Table 5 F1 score (%) on test set for various features in Table 3 and combinations fed to non-neural ML algorithms (Section 4.1) for indirectness detection on reciprocal peer-tutoring dataset.

	Reciprocal Peer-Tutoring Corpus (Indirectness Detection)	Wikipedia (Uncertainty Detection)	BioScope (Uncertainty Detection)
Att_Inp CNN	**62.03**	65.13*	84.99*
Att_Conv CNN	61.4	**66.49***	84.69*
Pre-trained W2V + Stacked LSTM	61.15	66.07	82.62
Pre-trained W2V + Stacked LSTM + Visual	61.35	-	-
Unigram + Stacked LSTM	56.5	43.71	73.03
Unigram + Stacked LSTM + Visual	57.11	-	-
SVM on Bag-of-Words	58.28	60.2*	**85.2***

Table 6 F1 score (%) for the various neural models compared with SVM approach for two different tasks on different datasets. Results marked with * have been taken from previous literature as explained in Section 5.

Attention based CNN model by [1] gave state-of-the-art results on the shared task datasets. We try these approaches for our task of indirectness detection, along with various other neural architectures (Section 4.2). To further establish the effectiveness of our stacked LSTMs approach (Section 4.2.1), we see its performance on the uncertainty detection shared task datasets as well (with the tuned hyperparameter settings reported in Table 4).

We show results on the test set for those variants of the neural and non-neural models which performed the best on validation set in Tables 5 and 6, and observe -

- SVM outperforms Logistic Regression for almost every feature representation by a significant margin (Table 5). Adding visual features does not seem to offer much improvement in terms of results, pointing to the need of looking at different nonverbal behaviors, or fusing them with verbal features in a different way.
- Using bag-of-words or n-grams obtained from our peer tutoring dataset gives better performance than using word2vec models pre-trained on other, much larger datasets (Table 5). This may indicate reliance of tasks like indirectness detection on the specific domain, as hinted by [36]. Among the pre-trained word2vec models, Twitter Word2Vec gave the best performance, and many utterances in our corpora do share the short length and informal nature of Twitter tweets.
- Using Pre-trained Word2Vec + Stacked LSTMs as well as Attention based CNN performs better on our dataset. These neural models outperform SVM by roughly 3-4% F1 score (second column of Table 6). The observation holds true for the uncertainty detection on the Wikipedia dataset as well (a performance gain of about 6%). This indicates that neural models constitute a viable approach for indirectness detection as well as uncertainty detection. and reinforces the importance of capturing context as well as possible for the task at hand. For BioScope corpus, however, SVM on bag-of-words based features performs the best, which may indicate a greater reliance on certain keywords indicating uncertainty compared to capturing the whole context of the sentence for that dataset. This is backed by another observation - using unigrams as input to stacked LSTMs resulted in massive performance downgrades for Wikipedia and our RPT datasets, but not as much reduction for the BioScope corpus. Note that 'context' here is in relation to capturing the overall meaning or content of the single utterance, and not going beyond one utterance.

- The best results we obtained on automatic detection of indirectness in peer-tutoring is 62.03% F1 score (Table 6), and the neural methods performed well for uncertainty detection in other domains as well.

6 Conclusion and Future Work

Indirectness is often used to mitigate face threat in conversations in various settings like business negotiations, doctor-patient discourse, counseling, conference talks, and in tutoring. Detecting indirectness can help virtual conversational agents and spoken dialogue systems respond to the user in a more appropriate manner. This may entail being more effective in their task goals (business deals, medical advice or tutoring instructions) or in managing their interpersonal relationship with the user (i.e. mitigating face threat and building trust). To achieve this, we train our models on a corpus of peer tutoring dialogues, using nonverbal behaviors in conjunction with the text of the transcripts. We hope that insights from our experiments will help inform the design of future spoken dialogue systems that can automatically detect indirect delivery from users. For example, the Twitter-like nature of collaborative educational conversations can be exploited like we did using a word2vec model pre-trained on tweets. Neural approaches (like stacked LSTMs and attention based CNNs) outperform non-neural approaches (like SVM), which hints at the need to capture the whole context (the overall meaning of the utterance and not just certain keywords) since indirectness occurs in various ways in dialogue. Such observations may be useful for spoken dialogue systems, regardless of the domain.

We intend for this work to be the first step towards automatic detection of indirect language using multi-modal data. For future work, we plan to leverage more visual behaviors by studying how various nonverbal behaviors inform the use of indirectness via dedicated experiments to annotate more behaviors backed by literature, such as head tilts [20] and laughter [45] (we currently have head nod and smile). We also aim to use acoustic or paralinguistic features to create a fully multimodal system for indirectness detection. Another line of work we hope to explore is properly incorporating our findings into an Intelligent Tutoring Agent or a general-purpose socially-aware spoken dialogue system [28] that can detect and use indirectness strategically. As indirectness is ubiquitous in interpersonal communication, incorporating its detection in spoken dialogue systems may ultimately lead to more natural, human-like interactions with users.

References

1. Adel, H., Schütze, H.: Exploring different dimensions of attention for uncertainty detection. In: EACL 2017, European Chapter of the Association for Computational Linguistics, Valencia, Spain, April 3 - April 7, 2017 (2017)

2. Aubakirova, M., Bansal, M.: Interpreting neural networks to improve politeness comprehension. arXiv preprint arXiv:1610.02683 (2016)
3. Blum-Kulka, S.: Indirectness and politeness in requests: Same or different? Journal of Pragmatics **11**(2), 131–146 (1987)
4. Brown, P., Levinson, S.C.: Politeness: Some universals in language usage, vol. 4. Cambridge University Press (1987)
5. Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014)
6. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. Journal of Machine Learning Research **12**(Aug), 2493–2537 (2011)
7. Cutting, J.: Vague language explored. Springer (2007)
8. Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., Potts, C.: A computational approach to politeness with application to social factors. arXiv preprint arXiv:1306.6078 (2013)
9. Dral, J., Heylen, D., et al.: Detecting uncertainty in spoken dialogues: an explorative research to the automatic detection of a speakers' uncertainty by using prosodic markers (2008)
10. Enríquez, F., Troyano, J.A., López-Solaz, T.: An approach to the use of word embeddings in an opinion classification task. Expert Systems with Applications **66**, 1–6 (2016)
11. Farkas, R., Vincze, V., Móra, G., Csirik, J., Szarvas, G.: The conll-2010 shared task: learning to detect hedges and their scope in natural language text. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task, pp. 1–12. Association for Computational Linguistics (2010)
12. Fraser, B.: Pragmatic competence: The case of hedging. New Approaches to hedging **1534** (2010)
13. Georgescul, M.: A hedgehop over a max-margin framework using hedge cues. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task, pp. 26–31. Association for Computational Linguistics (2010)
14. Godin, F., Vandersmissen, B., De Neve, W., Van de Walle, R.: Multimedia lab@ acl w-nut ner shared task: named entity recognition for twitter microposts using distributed word representations. ACL-IJCNLP **2015**, 146–153 (2015)
15. Graves, A., Jaitly, N.: Towards end-to-end speech recognition with recurrent neural networks. In: ICML, vol. 14, pp. 1764–1772 (2014)
16. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)
17. Hoque, M., Morency, L.P., Picard, R.W.: Are you friendly or just polite?–analysis of smiles in spontaneous face-to-face interactions. In: International Conference on Affective Computing and Intelligent Interaction, pp. 135–144. Springer (2011)
18. Jin, J., Bridges, S.M.: Educational technologies in problem-based learning in health sciences education: a systematic review. Journal of Medical Internet Research **16**(12) (2014)
19. Johnson, W.L., Rizzo, P.: Politeness in tutoring dialogs:run the factory, thats what id do. In: Intelligent Tutoring Systems, pp. 206–243. Springer (2004)
20. Kang, S.H., Gratch, J., Sidner, C., Artstein, R., Huang, L., Morency, L.P.: Towards building a virtual counselor: modeling nonverbal behavior during intimate self-disclosure. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1, pp. 63–70. International Foundation for Autonomous Agents and Multiagent Systems (2012)
21. Khanpour, H., Guntakandla, N., Nielsen, R.: Dialogue act classification in domain-independent conversations using a deep recurrent neural network. In: COLING, pp. 2012–2021 (2016)
22. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
23. Krahmer, E., Swerts, M.: How children and adults produce and perceive uncertainty in audiovisual speech. Language and speech **48**(1), 29–53 (2005)

24. Kulik, J.A., Fletcher, J.: Effectiveness of intelligent tutoring systems: a meta-analytic review. Review of Educational Research **86**(1), 42–78 (2016)
25. Lee, J.Y., Dernoncourt, F.: Sequential short-text classification with recurrent and convolutional neural networks. arXiv preprint arXiv:1603.03827 (2016)
26. Liscombe, J., Hirschberg, J., Venditti, J.J.: Detecting certainness in spoken tutorial dialogues. In: INTERSPEECH, pp. 1837–1840 (2005)
27. Madaio, M., Cassell, J., Ogan, A.: The impact of peer tutors use of indirect feedback and instructions. Philadelphia, PA: International Society of the Learning Sciences. (2017)
28. Matsuyama, Y., Bhardwaj, A., Zhao, R., Romeo, O., Akoju, S., Cassell, J.: Socially-aware animated intelligent personal assistant agent. In: SIGDIAL Conference, pp. 224–227 (2016)
29. McQuiddy, I.W.: Some conventional aspects of indirectness in conversation. Ph.D. thesis, University of Texas at Austin (1986)
30. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
31. Neary-Sundquist, C.: The use of hedges in the speech of esl learners. Elia (13), 149 (2013)
32. Palinscar, A.S., Brown, A.L.: Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction **1**(2), 117–175 (1984)
33. Park, E., Han, X., Berg, T.L., Berg, A.C.: Combining multiple sources of knowledge in deep cnns for action recognition. In: Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, pp. 1–8. IEEE (2016)
34. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
35. Person, N.K., Kreuz, R.J., Zwaan, R.A., Graesser, A.C.: Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. Cognition and Instruction **13**(2), 161–188 (1995)
36. Prokofieva, A., Hirschberg, J.: Hedging and speaker commitment. In: 5th Intl. Workshop on Emotion, Social Signals, Sentiment & Linked Open Data, Reykjavik, Iceland (2014)
37. Reynolds, M.: Chatbots learn how to drive a hard bargain (2017)
38. Rowland, T.: well maybe not exactly, but its around fifty basically?: Vague language in mathematics classrooms. In: Vague language explored, pp. 79–96. Springer (2007)
39. Rubinstein, R.: The cross-entropy method for combinatorial and continuous optimization. Methodology and Computing in Applied Probability **1**(2), 127–190 (1999)
40. Rundquist, S.: Indirectness in conversation: flouting grices maxims at dinner. In: Annual Meeting of the Berkeley Linguistics Society, vol. 16, pp. 509–518 (1990)
41. Searle, J.R.: Indirect speech acts. na (1975)
42. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research **15**(1), 1929–1958 (2014)
43. Täckström, O., Velupillai, S., Hassel, M., Eriksson, G., Dalianis, H., Karlgren, J.: Uncertainty detection as approximate max-margin sequence labelling. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task, pp. 84–91. Association for Computational Linguistics (2010)
44. Trees, A.R., Manusov, V.: Managing face concerns in criticism integrating nonverbal behaviors as a dimension of politeness in female friendship dyads. Human Communication Research **24**(4), 564–583 (1998)
45. Warner-Garcia, S.: Laughing when nothings funny: The pragmatic use of coping laughter in the negotiation of conversational disagreement. Pragmatics **24**(1), 157–180 (2014)
46. Yin, W., Schütze, H., Xiang, B., Zhou, B.: Abcnn: Attention-based convolutional neural network for modeling sentence pairs. arXiv preprint arXiv:1512.05193 (2015)
47. Yu, H., Gui, L., Madaio, M., Ogan, A., Cassell, J., Morency, L.P.: Temporally selective attention model for social and affective state recognition in multimedia content (2017)
48. Zhang, G.: The impact of touchy topics on vague language use. Journal of Asian Pacific Communication **23**(1), 87–118 (2013)

49. Zhao, D., Nie, J.: Vague language in business negotiation-from a pragmatics perspective. Theory and Practice in Language Studies **5**(6), 1257 (2015)
50. Zhao, R., Sinha, T., Black, A.W., Cassell, J.: Automatic recognition of conversational strategies in the service of a socially-aware dialog system. In: 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, p. 381 (2016)
51. Zhou, C., Sun, C., Liu, Z., Lau, F.: A c-lstm neural network for text classification. arXiv preprint arXiv:1511.08630 (2015)