Decoupling Strategy and Generation in Negotiation Dialogues

He He and **Derek Chen** and **Anusha Balakrishnan** and **Percy Liang**
Computer Science Department, Stanford University
{hehe,derekchen14,anusha,pliang}@cs.stanford.edu

Abstract

We consider negotiation settings in which two agents use natural language to bargain on goods. Agents need to decide on both high-level strategy (e.g., proposing $50) and the execution of that strategy (e.g., generating *"The bike is brand new. Selling for just $50!"*). Recent work on negotiation trains neural models, but their end-to-end nature makes it hard to control their strategy, and reinforcement learning tends to lead to degenerate solutions. In this paper, we propose a modular approach based on coarse dialogue acts (e.g., propose(price=50)) that decouples strategy and generation. We show that we can flexibly set the strategy using supervised learning, reinforcement learning, or domain-specific knowledge without degeneracy, while our retrieval-based generation can maintain context-awareness and produce diverse utterances. We test our approach on the recently proposed DEALORNODEAL game, and we also collect a richer dataset based on real items on Craigslist. Human evaluation shows that our systems achieve higher task success rate and more human-like negotiation behavior than previous approaches.

1 Introduction

A good negotiator needs to decide on the *strategy* for achieving a certain goal (e.g., proposing $6000) and the realization of that strategy via *generation* of natural language (e.g., *"I really need a car so I can go to work, but all I have is 6000, any more and I won't be able to feed my children."*).

Most past work in NLP on negotiation focuses on strategy (dialogue management) with either no natural language (Cuayáhuitl et al., 2015; Cao et al., 2018) or canned responses (Keizer et al., 2017; Traum et al., 2008). Recently, end-to-end neural models (Lewis et al., 2017; He et al., 2017) are used to simultaneously learn dialogue strategy and language realization from human-human dialogues, following the trend of using neural network models on both goal-oriented dialogue (Wen et al., 2017a; Dhingra et al., 2017) and open-domain dialogue (Sordoni et al., 2015; Li et al., 2017; Lowe et al., 2017). However, these models have two problems: (i) it is hard to control and interpret the strategies, and (ii) directly optimizing the agent's goal through reinforcement learning often leads to degenerate solutions where the utterances become ungrammatical (Lewis et al., 2017) or repetitive (Li et al., 2016).

To alleviate these problems, our key idea is to decouple strategy and generation, which gives us control over the strategy such that we can achieve different negotiation goals (e.g., maximizing utility, achieving a fair deal) with the same language generator. Our framework consists of three components shown in Figure 1: First, the parser identifies keywords and entities to map each utterance to a *coarse dialogue act* capturing the high-level strategic move. Then, the dialogue manager chooses a responding dialogue act based on a sequence-to-sequence model over coarse dialogue acts learned from parsed training dialogues. Finally, the generator produces an utterance given the dialogue act and the utterance history.

Our framework follows that of traditional goal-oriented dialogue systems (Young et al., 2013), with one important difference: coarse dialogue acts are not intended to and cannot capture the full meaning of an utterance. As negotiation dialogues are fairly open-ended, the generator needs to depend on the full utterance history. For example, consider the first turn in Figure 1. We cannot generate a response given only the dialogue act inform; we must also look at the previous question. However, we still optimize the dialogue manager in the coarse dialogue act space using supervised learning, reinforcement learning, or domain-

Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2333–2343
Brussels, Belgium, October 31 - November 4, 2018.

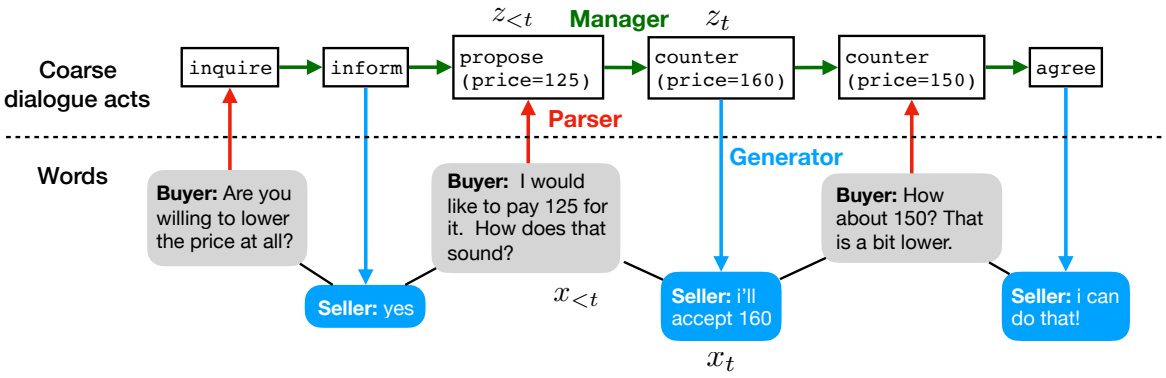

Figure 1: Our modular framework consists of three components similar to traditional goal-oriented dialogue systems. (1) The parser maps received utterances to coarse dialogue acts (an intent and its arguments) that capture the high-level dialogue flow. (2) The manager generates the next coarse dialogue act z_t conditioned on past dialogue acts $z_{<t}$. (3) The generator then produces a response conditioned on both the predicted coarse dialogue act z_t and the dialogue history $x_{<t}$. Importantly, unlike in traditional systems, coarse dialogue acts only capture the rough shape of a dialogue, not the full meaning of its utterances, e.g., inform does not specify the answer to the question.

specific knowledge.

Existing human-human negotiation datasets are grounded in closed-domain games with a fixed set of objects such as Settlers of Catan (lumber, coal, brick, wheat, and sheep) (Afantenos et al., 2012; Asher et al., 2016) or item division (book, hat, and ball) (DeVault et al., 2015; Lewis et al., 2017). These objects lack the richness of the real world. To study human negotiation in more open-ended settings that involve real goods, we scraped postings of items for sale from craigslist.org as our negotiation scenario. By hiring workers on Amazon Mechanical Turk (AMT) to play the role of buyers and sellers, we collected a new dataset (CRAIGSLISTBARGAIN) of negotiation dialogues.[1] Compared to existing datasets, our more realistic scenario invites richer negotiation behavior involving open-ended aspects such as cheap talk or side offers.

We evaluate two families of systems modeling coarse dialogue acts and words respectively, which are optimized by supervised learning, reinforcement learning, or domain knowledge. Each system is evaluated on our new CRAIGSLISTBARGAIN dataset and the DEALORNODEAL dataset of Lewis et al. (2017) by asking AMT workers to chat with the system in an A/B testing setting. We focus on two metrics: task-specific scores (e.g., utility) and human-likeness. We show that reinforcement learning on coarse dialogue acts avoids degenerate solutions, which was a problem in Li et al. (2016); Lewis et al. (2017). Our modular model maintains reasonable human-like behavior while still optimizes the objective. Furthermore, we find that models trained over coarse dialogue acts are stronger negotiators (even with only supervised learning) and produce more diverse utterances than models trained over words. Finally, the interpretability of coarse dialogue acts allows system developers to combine the learned dialogue policy with hand-coded rules, thus imposing stronger control over the desired strategy.

2 Craigslist Negotiation Dataset

Previous negotiation datasets were collected in the context of games. For example, Asher et al. (2016) collected chat logs from online Settlers of Catan. Lewis et al. (2017) asked two people to divide a set of hats, books, and balls. While such games are convenient for grounding and evaluation, it restricts the dialogue domain and the richness of the language. Most utterances are direct offers such as *"has anyone got wood for me?"* and *"I want the ball."*, whereas real-world negotiation would involve more information gathering and persuasion.

To encourage more open-ended, realistic negotiation, we propose the CRAIGSLISTBARGAIN task. Two agents are assigned the role of a buyer and a seller; they are asked to negotiate the price of an item for sale on Craigslist given a description and photos. As with the real platform, the listing price is shown to both agents. We addition-

[1] Available at https://stanfordnlp.github.io/cocoa.

JVC HD-ILA 1080P 70 Inch TV

Tv is approximately 10 years old. Just installed new lamp. There are 2 HDMI inputs. Works and looks like new.
Listing price: $275
Buyer's target price: $192

Agent	Utterance	Dialogue Act
Buyer	Hello do you still have the TV?	greet
Seller	Hello, yes the TV is still available	greet
Buyer	What condition is it in? Any scratches or problems? I see it recently got repaired	inquire
Seller	It is in great condition and works like a champ! I just installed a new lamp in it. There aren't any scratches or problems.	inform
Buyer	All right. Well I think 275 is a little high for a 10 year old TV. Can you lower the price some? How about 150?	propose(150)
Seller	I am willing to lower the price, but $150 is a little too low. How about $245 and if you are not too far from me, I will deliver it to you for free?	counter(245)
Buyer	It's still 10 years old and the technology is much older. Will you do 225 and you deliver it. How's that sound?	counter(225)
Seller	Okay, that sounds like a deal!	agree
Buyer	Great thanks!	agree
Seller	OFFER $225.0	offer(225)
Buyer	ACCEPT	accept

Table 1: Example dialogue between two people negotiating the price of a used TV.

ally suggest a private price to the buyer as a target. Agents chat freely in alternating turns. Either agent can enter an offer price at any time, which can be accepted or rejected by the partner. Agents also have the option to quit, in which case the task is completed with no agreement.

To generate the negotiation scenarios, we scraped postings on `sfbay.craigslist.org` from the 6 most popular categories (housing, furniture, cars, bikes, phones, and electronics). Each posting produces three scenarios with the buyer's target prices at 0.5x, 0.7x and 0.9x of the listing price. Statistics of the scenarios are shown in Table 2.

We collected 6682 human-human dialogues on AMT using the interface shown in Appendix A Figure 2. The dataset statistics in Table 3 show that CRAIGSLISTBARGAIN has longer dialogues and more diverse utterances compared to prior datasets. Furthermore, workers were encouraged to embellish the item and negotiate side offers such as free delivery or pick-up. This highly relatable scenario leads to richer dialogues such as the one shown in Table 1. We also observed various persuasion techniques listed in Table 4 such as embellishment, side offers, and appeals to sympathy.

# of unique postings	1402
% with images	80.8
Avg # of tokens per description	42.6
Avg # of tokens per title	33.8
Vocab size	12872

Table 2: Statistics of CRAIGSLISTBARGAIN scenarios.

	CB	DN	SoC
# of dialogues	6682	5808	1081
Avg # of turns	9.2	6.6	8.5
Avg # of tokens per turn	15.5	7.6	4.2
Vocab size	13928	2719	4921
Vocab size (excl. numbers)	11799	2623	4735

Table 3: Comparison of dataset statistics of CRAIGSLISTBARGAIN (CB), DEALORNODEAL (DN), and SETTLERSOFCATAN (SoC). CRAIGSLISTBARGAIN contains longer, more diverse dialogues on average.

3 Approach

3.1 Motivation

While end-to-end neural models have made promising progress in dialogue systems (Wen et al., 2017a; Dhingra et al., 2017), we find they struggle to simultaneously learn the strategy and the rich utterances necessary to succeed in the CRAIGSLISTBARGAIN domain, e.g., Table 8(a) shows a typical dialogue between a human and a sequence-to-sequence-based bot, where the bot easily agrees. We wish to now separate negotiation strategy and language generation. Suppose the buyer says: *"All right. Well I think 275 is a little high for a 10 year old TV. Can you lower the price some? How about 150?"* We can capture the highest-order bit with a coarse dialogue act propose(price=150). Then, to generate the seller's response, the agent can first focus on this coarse

Phenomenon	Example
Embellishment	It is in great condition and **works like a champ! I just installed a new lamp in it.** There aren't any scratches or problems.
Cheap talk	How about i give you $20 and you keep the helmet. **its for my daughter for her job, she delivers lemonade**.
Side offers	**Throw in a couple of movies** with that DVD player, and you have yourself a deal.
Appeal to sympathy	I would love to have this for my mother, **she is very sick** and this would help her and with me taking care of her and having to take a leave from work I can't pay very much of it
World knowledge	For a **Beemer 5 series** in this condition, I really can't go that low.

Table 4: Rich negotiation language in our CRAIGSLISTBARGAIN dataset.

dialogue act rather than having to ingest the free-form text all at once. Once a counter price is decided, the rest is open-ended justification for the proposed price, e.g., emphasizing the quality of the TV despite its age.

Motivated by these observations, we now describe a modular framework that extracts coarse dialogue acts from utterances, learns to optimize strategy in the dialogue act space, and uses retrieval to fill in the open-ended parts conditioned on the full dialogue history.

3.2 Overview

Our goal is to build a dialogue agent that takes the dialogue history, i.e. a sequence of utterances $x_1, \ldots, x_{t-1}$ along with the dialogue scenario c (e.g., item description), and produces a distribution over the responding utterance x_t.

For each utterance x_t (e.g., *"I am willing to pay $15"*), we define a coarse dialogue act z_t (e.g., propose(price=15)); the coarse dialogue act serves as a logical skeleton which does not attempt to capture the full semantics of the utterance. Following the strategy of traditional goal-oriented dialogue systems (Young et al., 2013), we broadly define our model in terms of the following three modules:

1. A **parser** that (deterministically) maps an input utterance x_{t-1} into a coarse dialogue act z_{t-1} given the dialogue history $x_{<t}$ and $z_{<t}$, as well as the scenario c.
2. A **manager** that predicts the responding dialogue act z_t given past coarse dialogue acts $z_{<t}$ and the scenario c.
3. A **generator** that turns the coarse dialogue act z_t to a natural language response x_t given the full dialogue history $x_{<t}$.

Because coarse dialogue acts do not capture the full semantics, the parser and the generator maintains full access to the dialogue history. The main restriction is the manager examining the dialogue acts, which we show will reduce the risk of degeneracy during reinforcement learning Section 4.4. We now describe each module in detail (Figure 1).

3.3 Parser

Our framework is centered around the coarse dialogue act z, which consists of an intent and a set of arguments. For example, *"I am willing to pay $15"* is mapped to propose(price=15). The fact that our coarse dialogue acts do not intend to capture the full semantics of a sentence allows us to use a simple rule-based parser. It detects the intent and its arguments by regular expression matching and a few if-then rules. Our parser starts by detecting entities (e.g., prices, objects) and matching keyword patterns (e.g., *"go lower"*). These signals are checked against an ordered list of rules, where we choose the first matched intent in the case of multiple matches. An unknown act is output if no rule is triggered. The list of intent parsing rules used are shown in Table 5. Please refer to Appendix B for argument parsing based on entity detection.

3.4 Manager

The dialogue manager decides what action z_t the dialogue agent should take at each time step t given the sequence of past coarse dialogue acts $z_{<t}$ and the scenario c. Below, we describe three ways to learn the dialogue manager with increasing controllability: modeling human behavior in the training corpus (supervised learning), explicitly optimizing a reward function (reinforcement learning), and injecting hand-coded rules (hybrid policy).

Supervised learning. Given a parsed training corpus, each training example is a sequence of coarse dialogue acts over one dialogue, $z_1, \ldots, z_T$. We learn the transition probabilities

Generic Rules	
Intent	Matching Patterns
greet	*hi, hello, hey, hiya, howdy*
disagree	*no, not, n't, nothing, dont*
agree	not disagree and *ok, okay, great, perfect, deal, that works, i can do that*
insist	the same offer as the previous one is detected
inquire	starts with an interrogative word (e.g., *what, when, where*) or particle (e.g., *do, are*)
CRAIGSLISTBARGAIN Rules	
Intent	Matching Patterns
intro	greet or *how are you, interested*
propose	first price mention
vague-price	no price mention and *come down, highest, lowest, go higher/lower, too high/low*
counter	new price detected
inform	previous coarse dialogue act was inquire
DEALORNODEAL Rules	
Intent	Matching Patterns
propose	items and respective counts are detected

Table 5: Rules for intent detection in the parser.

$p_\theta(z_t \mid z_{<t}, c)$ by maximizing the likelihood of the training data.

We use a standard sequence-to-sequence model with attention. Each coarse dialogue act is represented as a sequence of tokens, i.e. an intent followed by each of its arguments, e.g., "offer 150". During the agent's listening turn, an LSTM encodes the received coarse dialogue act; during its speaking turn, another LSTM decodes the tokens in the coarse dialogue act. The hidden states are carried over the entire dialogue to provide full history.

The vocabulary of coarse dialogue acts is much smaller than the word vocabulary. For example, our implementation includes fewer than 10 intents and argument values are normalized and binned (see Section 4.2).

Reinforcement learning. Supervised learning aims to mimic the average human behavior, but sometimes we want to directly optimize for a particular dialogue goal. In reinforcement learning, we define a reward $R(z_{1:T})$ on the entire sequence of coarse dialogue acts. Specifically, we experiment with three reward functions:

- **Utility** is the objective of a self-interested agent. For CRAIGSLISTBARGAIN, we set the utility function to be a linear function of the final price, such that the buyer has a utility of 1 at their target price, the seller has a utility of 1 at the listing price, and both agents have a utility of zero at the midpoint of the listing price and the buyer's target price, making it a zero-sum game. For DEALORNODEAL, utility is the total value of objects given to the agent.
- **Fairness** aims to achieve equal outcome for both agents, i.e. the difference between two agents' utilities.
- **Length** is the number of utterances in a dialogue, thus encourages agents to chat as long as possible.

The reward is -1 if no agreement is reached.

We use policy gradient (Williams, 1992) for optimization. Given a sampled trajectory $z_{1:T}$ and the final reward r, let a_i be the i-th generated token (i.e. "action" taken by the policy) along the trajectory. We update the parameters θ by

$$\theta \leftarrow \theta - \eta \sum_i \nabla_\theta \log p_\theta(a_i \mid a_{<i}, c)(r - b) \quad (1)$$

where η is the learning rate and b is a baseline estimated by the average return so far for variance reduction.

Hybrid policy. Given the interpretable coarse dialogue acts, a simple option is to write a rule-based manager with domain knowledge, e.g., if $z_{t-1} =$ greet, then $z_t =$ greet. We combine these rules with a learned manager to fine-tune the dialogue policy. Specifically, the dialogue manager predicts the intent from a learned sequence model but fills in the arguments (e.g., price) using rules. For example, given a predicted intent propose, we can set the price to be the average of the buyer's and seller's current proposals (a split-the-difference strategy).

3.5 Generator

We use retrieval-based generation to condition on both the coarse dialogue act and the dialogue history. Each candidate in our database for retrieval is a tuple of an utterance x_t and its dialogue context x_{t-1}, represented by both templates and coarse dialogue acts. i.e. $(d(x_{t-1}), z_{t-1}, d(x_t), z_t)$, where d is the template extractor. Specifically, given a parsed training set, each utterance is converted to a template by delexicalizing arguments in its coarse dialogue act. For example, *"How about $150?"*

becomes "How about [price]?", where [price] is a placeholder to be filled in at generation time.

At test time, given z_t from the dialogue manager, the generator first retrieves candidates with the same intent as z_t and z_{t-1}. Next, candidates are ranked by similarity between their context templates and the current dialogue context. Specifically, we represent the context $d(x_{t-1})$ as a TF-IDF weighted bag-of-words vector and similarity is computed by a dot product of two context vectors. To encourage diversity, the generator samples an utterance from the top K candidates according to the distribution given by a trigram language model estimated on the training data.

4 Experiments

4.1 Tasks

We test our approach on two negotiation tasks. **CRAIGSLISTBARGAIN** (Section 2) asks a buyer and a seller to negotiate the price of an item for sale given its Craigslist post. **DEALORNODEAL** (Lewis et al., 2017) asks two agents to divide a set of items given their private utility functions.

4.2 Models

We compare two families of models: end-to-end neural models that directly map the input dialogue context to a sequence of output words, and our modular models that use coarse dialogue acts as the intermediate representation.

We start by training the word-based model and the act-based model with supervised learning (SL).

- **SL(word)**: a sequence-to-sequence model with attention over previous utterances and the scenario, both embedded as a continuous Bag-of-Words;
- **SL(act)**: our model described in Section 3 with a rule-based parser, a learned neural dialogue manager, and a retrieval-based generator.

To handle the large range of argument values (prices) in CRAIGSLISTBARGAIN for act-based models, we normalize the prices such that an agent's target price is 1 and the bottomline price is 0. For the buyer, the target is given and the bottomline is the listing price. For the seller, the target is the listing price and the bottomline is set to 0.7x of the listing price. The prices are then binned according to their approximate values with two digits after the decimal point.

Model	z	Parser	Manager	Generator
SL/RL(word)	vector	learned	learned	generative
SL/RL(act)	logical	rules	learned	retrieval
SL(act)+rule	logical	rules	hybrid	retrieval

Table 6: Comparison of different implementation of the core modules in our framework.

Next, given the pretrained SL models, we fine-tune them with the three reward functions (Section 3.4), producing $\mathbf{RL_{utility}}$, $\mathbf{RL_{fairness}}$, and $\mathbf{RL_{length}}$.

In addition, we compare with the hybrid model, **SL(act)+rule**. It predicts the next intent using a trigram language model learned over intent sequences in the training data, and fills in the arguments with hand-coded rules. For CRAIGSLISTBARGAIN, the only argument is the price. The agent always splits the difference when making counter proposals, rejects an offer if it is worse than its bottomline and accepts otherwise. For DEALORNODEAL, the agent maintains an estimate of the partner's private utility function. In case of disagreement, it gives up the item with the lowest value of (own utility − partner utility) and takes an item of estimated zero utility to the partner. The agent agrees whenever a proposal is better than the last one or its predefined target. A high-level comparison of all models is shown in Table 6.

4.3 Training Details

CRAIGSLISTBARGAIN For SL(word), we use a sequence-to-sequence model with attention over 3 previous utterances and the negotiation scenario (embedded as a continuous Bag-of-Words). For both SL(word) and SL(act), we use 300-dimensional word vectors initialized by pretrained GloVe word vectors (Pennington et al., 2014), and a two-layer LSTM with 300 hidden units for both the encoder and the decoder. Parameters are initialized by sampling from a uniform distribution between -0.1 and 0.1. For optimization, we use AdaGrad (Duchi et al., 2010) with a learning rate of 0.01 and a mini-batch size of 128. We train the model for 20 epochs and choose the model with the lowest validation loss.

For RL, we first fit a partner model using supervised learning (e.g., SL(word)), then run RL

	CraigslistBargain					DealOrNoDeal				
	Hu	Ut	Fa	Ag	Len	Hu	Ut	Fa	Ag	Len
Human	4.3	-0.07	-0.14	0.91	10.2	4.6	5.5 vs. 5.3	-0.2	0.78	5.8
SL(word)	3.0	-0.32	-0.64	0.75	7.8	**3.8**	4.7 vs. 5.0	-0.3	0.70	5.0
SL(act)	3.3	0.06	-0.12	0.84	14.0	3.2	5.2 vs. 5.0	-0.2	0.67	7.0
SL(act)+rule	**3.6**	0.23	-0.46	0.75	11.4	**4.2**	5.2 vs. 5.2	0	0.72	8.0
$RL_{utility}$(word)	1.7	1.00	-2.00	0.31	2.5	1.7	2.9 vs. 1.8	-1.1	0.33	10.4
$RL_{utility}$(act)	**2.8**	1.00	-2.00	0.22	6.7	**2.8**	3.3 vs. 2.3	-1.0	0.38	9.5
$RL_{fairness}$(word)	1.8	-0.62	-1.24	0.75	9.4	3.2	5.7 vs. 5.9	-0.2	0.79	4.0
$RL_{fairness}$(act)	**3.0**	-0.28	-0.56	0.68	7.1	3.5	4.2 vs. 5.4	-1.2	0.77	7.6
RL_{length}(word)	1.9	-0.79	-1.58	0.85	13.8	1.6	3.4 vs. 2.9	-0.5	0.48	9.2
RL_{length}(act)	**3.0**	0.89	-1.78	0.40	11.8	**2.5**	2.5 vs. 3.1	-0.6	0.54	11.0

Table 7: Human evaluation results on human-likeness (Hu), agreement rate (Ag), and RL objectives, including agent utility (Ut), deal fairness (Fa), and dialogue length (Len). Results are grouped by the optimization objective. For each group of RL models, the column of the optimization objective is highlighted. For human-likeness, scores that are better than others in the same group with statistical significance ($p < 0.05$ given by paired t-tests) are in **bold**. Overall, with SL, all models are human-like, however, act-based models better matches human statistics across all metrics; with RL, word-based models becomes degenerate, whereas act-based models optimize the reward while maintaining human-likeness.

against it. One agent is updated by policy gradient and the partner model is fixed during training. We use a learning rate of 0.001 and train for 5000 episodes (dialogues). The model with the highest reward on the validation set is chosen.

DealOrNoDeal For act-based models, we use the same parameterization as CraigslistBargain. For word-based models, we use the implementation from Lewis et al. (2017).[2] Note that for fair comparison, we did not apply SL interleaving during RL training and rollouts during inference.

4.4 Human Evaluation

We evaluated each system on two metrics: task-specific scores (e.g., utility) and human-likeness. The scores tell us how well the system is playing the game, and human-likeness tells us whether the bot deviates from human behavior, presumably due to over-optimization.

We put up all 9 systems online and hired workers from AMT to chat with the bots. Each worker was randomly paired with one of the bots or another worker, so as to compare the bots with human performance under the same conditions. At the end of a chat, workers were asked the question *"Do you think your partner demonstrated reasonable human behavior?"*. They provided answers on a Likert scale from 1 (not at all) to 5 (definitely). Table 7 shows the human evaluation results on CraigslistBargain and DealOrNoDeal respectively. We also show example human-bot dialogues in Table 8 and Appendix C.

SL(act) learns more human-like behavior. We first compare performance of SL models over words and coarse dialogue acts. Both SL(word) and SL(act) achieved similar scores on human-likeness (no statistically significant difference). However, SL(word) better matched human statistics such as dialogue length and utility. For instance, SL(word) tended to produce short, generic utterances as shown in Table 8(a); they also agreed on a deal more quickly because utterances such as *"deal"* and *"I can do that"* are frequent in negotiation dialogues. This behavior is reflected by the shorter dialogue length and lower utility of SL(word) models.

RL(word) leads to degeneracy. On CraigslistBargain, all RL(word) models clearly have low scores on human-likeness in Table 7. They merely learned to repeat a few sentences: The three most frequent

[2] https://github.com/facebookresearch/end-to-end-negotiator

Title: Fitbit Flex + 2 Bands Activity Tracker **Listing Price:** $60
Description: Hi, I'm selling a Fitbit Activity Tracker. Track your day including steps, distance, calories burned, active minutes. Track your night including hours slept, times to woken up, sleep quality, Wake with a silent vibrating alarm. You can set a ...

(a) A: SL(word) (Buyer) B: Human (Seller)	(b) A: SL(act) (Buyer) B: Human (Seller)
A: **hi, i 'm interested in your fitbit.**	A: **how we doing today**
B: Great. It really is a good deal.	B: Hi!
A: **great.**	A: **yes, i really need one, what color is it?**
B: Do you have any questions about it?	B: It has a black band and a white band.
A: **i 'm interested in your fitbit. i 'm not sure i 'm not sure i 'm interested.**	A: **i see, so you bought an extra one by mistake?**
B: It has so many functions that are great. I particularly like the sleep tracking. Plus it is waterproof. How about I will deliver it to you and you give me 60 dollars?	B: No, it comes with 2 bands A: **i would like to make you a offer of 36** B: Ooooh....yeah, I can't. Sorry. It's wireless, has Bluetooth, extra bands, and is water resistant. I could do $50.
A: **i can do $52 if you come pick it up.**	A: **49?**
B: Well I am selling it. So I will deliver it for 60. Deal?	B: You know..I'd do $49.
A: **yes, i can do that.**	A: **thank you so much!**
B: OFFER($60) A: **ACCEPT**	B: OFFER($49) A: **ACCEPT**

(c) A: RL_{length}(word) (Buyer) B: Human (Seller)	(d) A: RL_{length}(act) (Buyer) B: Human (Seller)
A: **hello**	A: **hi, i saw your ad about the item.**
B: Hi how are you?	B: Okay great, i'm selling a Fitbit Flex plus 2 bands
A: **i am.**	A: **how old is the item?**
B: you are interested in the fitbit flex and the 2 trackers?	B: I've only had it for about a few months.
A: **it's in good condition condition.**	A: **but it does work correct?**
B: yes	B: yes it does work, it's in great condition.
A: **ok, that sounds good.**	A: **excellent. would you consider taking 36 for it?**
B: I am selling it at $60?	B: I'm selling for $60 but $36 would be way too low.
A: **it's in good condition condition.**	A: **how about 36?**
B: yes	B: no I cannot accept $36, I'm sorry
A: **ok, that sounds good.**	A: **i'm sorry. would you accept 36?**
B: $60	B: I am not going lower than $50
A: **i can go that low.**	A: **OFFER($36)** A: REJECT

Table 8: Example human-bot chats on CRAIGSLISTBARGAIN, where bot utterances are in bold. SL(word) produced generic responses, while SL(act) is more human-like. RL_{length}(word) devolved into degenerate behavior repeating itself while RL_{length}(act) maintained coherency. Only the first half of the item description and the RL_{length}(word) chat are shown due to space limit.

sentences of $RL_{utility}$(word), $RL_{fairness}$(word), and RL_{length}(word) account for 81.6%, 100% and 100% of all utterances. For example, $RL_{utility}$(word) almost always opened with *"i can pick it up"*, then offer its target price. RL_{length}(word) repeated generic sentences until the partner submitted a price. While they scored high on the reward being optimized, the conversations are unnatural.

On DEALORNODEAL, we have observed similar patterns. A general strategy learned by RL(word) was to pick an offer depending on its objective, then repeat the same utterance over and over again (e.g., *"i need the ball."*), resulting in low human-likeness scores. One exception is $RL_{fairness}$(word), since most of its offers were reasonable and agreed on immediately (it has the shorted dialogue length), the conversations are natural.

RL(act) optimizes different negotiation goals while being human-like. On both tasks, RL(act) models optimized their rewards while maintaining reasonable human-likeness scores. We now show that different models demonstrated different negotiation behavior. Two main strategies learned by RL_{length}(act) were to ask questions and to postpone offer submission. On CRAIGSLISTBARGAIN, when acting as a buyer, 42.4% of its utterances were questions, compared to 30.2% for other models. On both tasks, it tended to wait for the partner to submit an offer (even after a deal was agreed on), compared to RL_{margin}(act) which almost always submitted offers first. For $RL_{fairness}$(act), it aimed to agree on a price in the middle of the listing price and the buyer's target price for CRAIGSLISTBARGAIN. Since the buyer's target was hidden, when the agent was the seller, it tended to wait for the buyer to propose prices first. Similary, on DEALORN-

ODEAL it waited to hear the parter's offer and sometimes changed its offer afterwards, whereas the other models often insisted on one offer.

On both tasks, $RL_{utility}$(act) learned to insist on its offer and refuse to budge. This ended up frustrating many people, which is why it has a low agreement rate. The problem is that our human model is simply a SL model trained on human-human dialogues, which may not accurately reflects real human behavior during human-bot chat. For example, the SL model often agrees after a few turns of insistence on a proposal, whereas humans get annoyed if the partner is not willing to make compromises at all. However, by injecting domain knowledge to SL(act)+rule, e.g., making a small compromise is better than stubbornly being fixed on a single price, we were able to achieve high utility and human-likeness on both CRAIGSLISTBARGAIN and DEALORNODEAL.

5 Related Work and Discussion

Recent work has explored the space between goal-oriented dialogue and open-domain chit-chat through collaborative or competitive language games, such as collecting cards in a maze (Potts, 2012), finding a mutual friend (He et al., 2017), or splitting a set of items (DeVault et al., 2015; Lewis et al., 2017). Our CRAIGSLISTBARGAIN dialogue falls in this category, but exhibits richer and more diverse language than prior datasets. Our dataset calls for systems that can handle both strategic decision-making and open-ended text generation.

Traditional goal-oriented dialogue systems build a pipeline of modules (Young et al., 2013; Williams et al., 2016). Due to the laborious dialogue state design and annotation, recent work has been exploring ways to replace these modules with neural networks and end-to-end training while still having a logical backbone (Wen et al., 2017a; Bordes and Weston, 2017; He et al., 2017). Our work is closely related to the Hybrid Code Network (Williams et al., 2017), but the key difference is that Williams et al. (2017) uses a neural dialogue state, whereas we keep a structured, interpretable dialogue state which allows for stronger top-down control. Another line of work tackles this problem by introducing latent stochastic variables to model the dialogue state (Wen et al., 2017b; Zhao et al., 2017; Cao and Clark, 2017). While the latent discrete variable allows for post-hoc discovery of dialogue acts and increased utterance diversity, it does not provide controllability over the dialogue strategy.

Our work is also related to a large body of literature on dialogue policies in negotiation (English and Heeman, 2005; Efstathiou and Lemon, 2014; Hiraoka et al., 2015; Cao et al., 2018). These work mostly focus on learning good negotiation policies in a domain-specific action space, whereas our model operates in an open-ended space of natural language. An interesting future direction is to connect with game theory (Brams, 2003) for complex multi-issue bargaining. Another direction is learning to generate persuasive utterances, e.g., through framing (Takuya et al., 2014) or accounting for the social and cultural context (Elnaz et al., 2012).

To conclude, we have introduced CRAIGSLISTBARGAIN, a rich dataset of human-human negotiation dialogues. We have also presented a modular approach based on coarse dialogue acts that models a rough strategic backbone as well allowing for open-ended generation. We hope this work will spur more research in hybrid approaches that can work in open-ended, goal-oriented settings.

Acknowledgments. This work is supported by DARPA Communicating with Computers (CwC) program under ARO prime contract no. W911NF-15-1-0462. We thank members of the Stanford NLP group for insightful discussion and the anonymous reviewers for constructive feedback.

Reproducibility. All code, data, and experiments for this paper are available on the CodaLab platform: https://worksheets.codalab.org/worksheets/0x453913e76b65495d8b9730d41c7e0a0c/.

References

S. Afantenos, N. Asher, F. Benamara, A. Cadilhac, C. Dégremont, P. Denis, M. Guhe, S. Keizer, A. Lascarides, O. Lemon, et al. 2012. Modelling strategic conversation: Model, annotation design and corpus. In *Proceedings of SemDial 2012: Workshop on the Semantics and Pragmatics of Dialogue*, pages 167–168.

N. Asher, J. Hunter, M. Morey, F. Benamara, and S. Afantenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus. In *Language Resources and Evaluation Conference (LREC)*.

A. Bordes and J. Weston. 2017. Learning end-to-end goal-oriented dialog. In *International Conference on Learning Representations (ICLR)*.

S. J. Brams. 2003. *Negotiation Games: Applying Game Theory to Bargaining and Arbitration*. Psychology Press.

K. Cao and S. Clark. 2017. Latent variable dialogue models and their diversity. In *European Association for Computational Linguistics (EACL)*.

K. Cao, A. Lazaridou, M. Lanctot, J. Z. Leibo, K. Tuyls, and S. Clark. 2018. Emergent communication through negotiation. In *International Conference on Learning Representations (ICLR)*.

H. Cuayáhuitl, S. Keizer, and O. Lemon. 2015. Strategic dialogue management via deep reinforcement learning. In *Advances in Neural Information Processing Systems (NIPS)*.

D. DeVault, J. Mell, and J. Gratch. 2015. Toward natural turn-taking in a virtual human negotiation agent. In *Association for the Advancement of Artificial Intelligence (AAAI)*.

B. Dhingra, L. Li, X. Li, J. Gao, Y. Chen, F. Ahmed, and L. Deng. 2017. End-to-end reinforcement learning of dialogue agents for information access. In *Association for Computational Linguistics (ACL)*.

J. Duchi, E. Hazan, and Y. Singer. 2010. Adaptive subgradient methods for online learning and stochastic optimization. In *Conference on Learning Theory (COLT)*.

I. Efstathiou and O. Lemon. 2014. Learning non-cooperative dialogue behaviours. In *Special Interest Group on Discourse and Dialogue (SIGDIAL)*.

N. Elnaz, G. Kallirroi, and T. David. 2012. A cultural decision-making model for negotiation based on inverse reinforcement learning. In *The Annual Meeting of the Cognitive Science Society*.

M. S. English and P. A. Heeman. 2005. Learning mixed initiative dialog strategies by using reinforcement learning on both conversants. In *Empirical Methods in Natural Language Processing (EMNLP)*.

H. He, A. Balakrishnan, M. Eric, and P. Liang. 2017. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *Association for Computational Linguistics (ACL)*, pages 1766–1776.

T. Hiraoka, K. Georgila, E. Nouri, and D. Traum. 2015. Reinforcement learning in multi-party trading dialog. In *Special Interest Group on Discourse and Dialogue (SIGDIAL)*.

S. Keizer, M. Guhe, H. Cuayahuitl, I. Efstathiou, K. Engelbrecht, M. Dobre, A. Lascarides, and O. Lemon. 2017. Evaluating persuasion strategies and deep reinforcement learning methods for negotiation dialogue agents. In *European Association for Computational Linguistics (EACL)*.

M. Lewis, D. Yarats, Y. N. Dauphin, D. Parikh, and D. Batra. 2017. Deal or no deal? end-to-end learning for negotiation dialogues. In *Empirical Methods in Natural Language Processing (EMNLP)*.

J. Li, W. Monroe, A. Ritter, D. Jurafsky, M. Galley, and J. Gao. 2016. Deep reinforcement learning for dialogue generation. In *Empirical Methods in Natural Language Processing (EMNLP)*.

J. Li, W. Monroe, T. Shi, A. Ritter, and D. Jurafsky. 2017. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*.

R. T. Lowe, N. Pow, I. Serban, L. Charlin, C. Liu, and J. Pineau. 2017. Training end-to-end dialogue systems with the ubuntu dialogue corpus. *Dialogue and Discourse*, 8.

J. Pennington, R. Socher, and C. D. Manning. 2014. GloVe: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

C. Potts. 2012. Goal-driven answers in the Cards dialogue corpus. In *Proceedings of the 30th West Coast Conference on Formal Linguistics*, pages 1–20.

A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J. Nie, J. Gao, and B. Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. In *North American Association for Computational Linguistics (NAACL)*.

H. Takuya, N. Graham, S. Sakriani, T. Tomoki, and N. Satoshi. 2014. Reinforcement learning of cooperative persuasive dialogue policies using framing. In *International Conference on Computational Linguistics (COLING)*.

D. Traum, S. C. Marsella, J. Gratch, J. Lee, and A. Hartholt. 2008. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *International Workshop on Intelligent Virtual Agents*, pages 117–130.

T. Wen, M. Gasic, N. Mrksic, L. M. Rojas-Barahona, P. Su, S. Ultes, D. Vandyke, and S. Young. 2017a. A network-based end-to-end trainable task-oriented dialogue system. In *European Association for Computational Linguistics (EACL)*, pages 438–449.

T. Wen, Y. Miao, P. Blunsom, and S. Young. 2017b. Latent intention dialogue models. In *International Conference on Machine Learning (ICML)*.

J. D. Williams, K. Asadi, and G. Zweig. 2017. Hybrid code networks: Practical and efficient end-to-end dialog control with supervised and reinforcement learning. In *Association for Computational Linguistics (ACL)*.

J. D. Williams, A. Raux, and M. Henderson. 2016. The dialog state tracking challenge series: A review. *Dialogue and Discourse*, 7.

R. J. Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256.

S. Young, M. Gašić, B. Thomson, and J. D. Williams. 2013. POMDP-based statistical spoken dialog systems: A review. In *Proceedings of the IEEE*, 5, pages 1160–1179.

T. Zhao, R. Zhao, and M. Eskenazi. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *Association for Computational Linguistics (ACL)*.