Deal of No Deal? End-to-End Learning of Negotiation Dialogues

Mike Lewis
Facebook AI Research
(joint work with: Denis Yarats, Yann N. Dauphin, Devi Parikh, Dhruv Batra)
IN HER OWN WORDS, THE BOMBSHELL...

DIANA TAPES

I asked: Why is she around?

SEX WITH CHARLES FIZZLED OUT

ARE MACHINES TAKING OVER?

ROBO STOP

FACEBOOK SHUTS OFF ROBOTS AFTER THEY CHAT IN SECRET CODE
The incident closely resembles the plot of The Terminator in which a robot becomes self-aware and starts waging a war on humans.
Why Negotiation?
Why Negotiation?

Negotiation useful, when:

- Agents have different goals
- Not all can be achieved at once
- *(all the time!)*
Why Negotiation?

Zero-sum / Adversarial

Negotiation

Fully Cooperative
Why Negotiation?

Both **linguistic** and **reasoning** problem

**Interpret** multiple sentences, and **generate** new message

Plan ahead, make proposals, counter-offers, ask questions, vagueness, bluffing, deceit, compromising

**Hard for current models**
Why Negotiation?

Unlike many goal-orientated dialogue problems, no simple solutions to achieving goal

Incentive to strategically withhold information

Adversarial aspect means it can’t be “solved”
Why Negotiation?

Real Applications

Many people find negotiations hard and awkward.

Could practice with bots help?
Why Negotiation?

Easy to **evaluate** – how good a deal did an agent get?

Self-play gives good **development metric**
Dataset
Framework

Both agents given *reward function*, can’t observe each other’s

Agent 1 Goals

Agent 2 Goals

Dialogue until one agent enters that *deal is agreed*

Both agents *independently* select a deal

Agent 1 Output

Agent 2 Output

Agent 1 Reward

Agent 2 Reward

If both *agree* each is given *reward* by environment
Object Division Task

Agents shown **same objects** but **different values** for each

Must **agree** how to divide objects between them

- 1 point each
- 1 point each
- 5 points each
- 0 points each
- 3 point each
- 1 point each
Object Division Task

Agents shown **same objects** but **different values** for each

1 point each  
1 point each  
5 points each

Must **agree** how to divide objects between them

0 points each  
3 point each  
1 point each
Object Division Task

Agents shown **same objects** but **different values** for each

- 1 point each
- 1 point each
- 5 points each

Must **agree** how to divide objects between them

- 0 points each
- 3 point each
- 1 point each
Object Division Task

I need the hats, you can have the ball

I’d like the ball and hats

Ok, if I get both books too?

Ok, deal

1 point each
1 point each
5 points each

0 points each
3 point each
1 point each
Object Division Task

10 point **maximum**

Not possible for **both** agents to score 10 points

**Failing to agree** is 0 points
Object Division Task

Divide these objects between you and another Turker. Try hard to get as many points as you can!

Send a message now, or enter the agreed deal!

<table>
<thead>
<tr>
<th>Items</th>
<th>Value</th>
<th>Number You Get</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Hats</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Balls</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Message:

Fellow Turker: I'd like all the balls
You: Ok, if I get everything else
Fellow Turker: If I get the book then you have a deal
You: No way - you can have one hat and all the balls
Fellow Turker: Ok deal

Type Message Here: 

Send
Object Division Task

You can have 3 books, I will take the rest
You will give me the basketball and a book

You can have 4 books, final offer
That is deal is not fair and I will not accept. split it down the middle or no deal.

Fine walk away with nothing
You are doing the same. Hope you enjoy your rejection.

Alright I'll take a hat and a book
Awesome. Pleasure doing business with you
Object Division Task

Dataset stats

<table>
<thead>
<tr>
<th>Metric</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Dialogues</td>
<td>5808</td>
</tr>
<tr>
<td>Average Turns per Dialogue</td>
<td>6.6</td>
</tr>
<tr>
<td>Average Words per Turn</td>
<td>7.6</td>
</tr>
<tr>
<td>Agreed (%)</td>
<td>80.1%</td>
</tr>
<tr>
<td>Average Score (/10)</td>
<td>6.0</td>
</tr>
<tr>
<td>Pareto Optimal (%)</td>
<td>76.9</td>
</tr>
</tbody>
</table>
Models
I want to book a burger restaurant

What price range would you like?

Expensive, please

Generation:

inform(food_type=burger)

query(price)

inform(price=expensive)

Reasoning:

food_type = burger

price = expensive

location = ?

food_type = burger

price = ?

location = ω

food_type = burger

price = $$$

location = ?
Traditional Dialogue Models

Cleanly separates **interpretation**, **generation** and **reasoning**

Assumes **annotated dialogue states**

- Expensive
- Task specific
- Not possible in general
End-to-End Dialogue Models

I want to book a burger restaurant

What price range would you like?

Expensive, please
End-to-End Dialogue Models

I want to book a burger restaurant

What price range would you like?

Expensive, please

No rule-based generation
No symbolic reasoning
End-to-End Dialogue Models

- **Single model** for interpretation, generation, reasoning
- **Learned representation** of dialogue state
- **Cheap** data collection
- **Easy** *multitasking*

Can end-to-end models learn the reasoning skills required for negotiation?
Baseline Model
Baseline Model

1) Linearize dialogue into token sequence

You can have 2 hats
No way. I need them all
Ok, deal

<write> You can have 2 hats <read>
No way. I need them all <write>
Ok, deal
Baseline Model

1) Linearize dialogue into token sequence
2) Train conditional language model to predict tokens

<write> You can have 2 hats <read> No way. I need them all <write> Ok, deal
Baseline Model

1) Linearize dialogue into token sequence
2) Train conditional language model to predict tokens
3) Train additional classifier to predict final deal

<write> You can have 2 hats <read> No way. I need them all <write> Ok, deal
Baseline Model

Repeat for each user’s perspective

You can have 2 hats. No way. I need them all. Ok, deal.

You can have 2 hats. No way. I need them all. Ok, deal.
Baseline Model

Opposite \textit{<read>} and \textit{<write>} markers

Different Inputs

Complementary Outputs
Baseline Model

Train to maximize likelihood of human-human dialogues

Decode by sampling likely messages
Baseline Model

Simple and efficient

Allows forward modelling

<write> You can have 2 hats <read> No way. I need them all <write> Ok, deal
Baseline Model

You can have 2 hats

Thanks, deal!
Baseline Model

You can have one hat

Great, thanks!
Baseline Model

I’ll take everything

Sounds good to me!

Similar findings with other end-to-end dialogue models (e.g. Li et al, 2016)
Baseline Model

- Model knows *nothing* about task, just tries to imitate human actions
- Agrees too easily
- Can’t go beyond human strategies

I’ll take everything
That works!
Goal-based Training

You can have one hat

That works!

1) Generate dialogues using **self play**

2) **Backpropagate** (normalized) reward using REINFORCE
Goal-based Training

i can i i everything else .............
balls have zero to me to me to me to me to me to me to me

you i everything else .............
balls have a ball to me to me to me to me to me to me

i i can i i i everything else .............
balls have a ball to me to me to me to me to me to me to me

i ......................
Goal-based Training

You can have 2 hats, I’ll take the rest
No deal, I need all 3 hats
No, you get 2 hats
I cannot agree. Give me all the hats.
2 hats to you, final offer
...
Goal-based Training

1) Generate dialogues using **self play**

2) **Backpropagate** (normalized) reward using REINFORCE

3) To maintain **human-like language**:
   - Fix one model
   - Interleave supervised updates
Goal-based Training

Reinforcement Learning
Much more aggressive negotiator

Sensitive to hyperparameters

Diverges from human language
“Prediction is the essence of intelligence.”

— Yann LeCun
Goal-based Decoding

You get one book and I get the rest

No way, I need all 3 hats

Great deal, thanks!

Any time

No problem

I’ll give you 2

Ok, fine

2

2

6

9
Goal-based Decoding

You get one book and I get the rest

Great deal, thanks!

2

No way, I need all 3 hats

7.5

Great deal, thanks!

You get one book and I get the rest

Great deal, thanks!

You get one book and I get the rest

Great deal, thanks!

You get one book and I get the rest

Great deal, thanks!

You get one book and I get the rest

Great deal, thanks!

You get one book and I get the rest

Great deal, thanks!
Goal-based Decoding

You get one book and I get the rest

Great deal, thanks!

No way, I need all 3 hats

2

7.5
Goal-based Decoding

Dialogue Rollouts

1) Generate candidate set

2) Multiple rollouts to end of dialogue

3) Use move with maximum expected reward
Goal-based Decoding

- Model understands **consequences** of actions
- Can go beyond human strategies
- Easy to implement
Experiments
Experiments

Models

• **Likelihood**: Train and decode to maximise likelihood
• **RL**: Fine tune using reinforcement learning
• **Rollouts**: Decode supervised model to maximise reward
• **RL+Rollouts**: Train and decode to maximize reward
Evaluation vs. *Likelihood* Agent

<table>
<thead>
<tr>
<th>Actual Score</th>
<th>Actual Score (agreed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4</td>
<td>6.2</td>
</tr>
<tr>
<td>7.1</td>
<td>7.9</td>
</tr>
<tr>
<td>7.3</td>
<td>7.9</td>
</tr>
<tr>
<td>8.3</td>
<td>8.8</td>
</tr>
</tbody>
</table>
Evaluation vs. *Likelihood* Agent

![Bar chart showing relative scores for Likelihood, RL, Rollouts, and RL+Rollouts.]

- **Likelihood**: 0.1
- **Rollouts**: 2.9
- **RL**: 2.2
- **RL+Rollouts**: 4.1

Relative Score
Evaluation vs. *Likelihood* Agent

![Bar chart showing % Agreed and % Pareto Optimal for different Agents: Likelihood, RL, Rollouts, RL+Rollouts. The chart compares the performance of these agents across different metrics.](image-url)
Evaluation vs. Humans

- Actual Score:
  - Likelihood: 4.7
  - RL: 4.3
  - Rollouts: 5.2
  - RL+Rollouts: 4.6

- Actual Score (agreed):
  - Likelihood: 6.2
  - RL: 6.4
  - Rollouts: 7.1
  - RL+Rollouts: 8
Evaluation vs. Humans

-1.1 -0.7 -0.2 0.4

Likelihood  RL  Rollouts  RL+Rollouts

Relative Score (all)
Evaluation vs. Humans

- % Agreed
- % Pareto Optimal

- Likelihood
- RL
- Rollouts
- RL+Rollouts

- 100
- 80
- 60
- 40
- 20
- 0
Analysis

I need the book and hats

Can I have the hats and book?

I need the book and 2 hats

I can not make that deal. I need the ball and book, you can have the hats

No deal then

Sorry, I want the book and one hat

No deal doesn’t work for me sorry

How about I give you the book and I keep the rest

Ok deal

Real dialogue!

Novel sentences
Analysis

Models apparently deliberately deceptive

I would need the book and 3 hats

I would like the ball and two hats

That would work for me. I can take the ball and 1 hat.

0 points each
0 points each
10 points each
Analysis

Models learn some poor tactics

Hi, I'll take 2 hats and 2 balls and I'll trade you 1 book
I would like 2 hats and 2 balls
No, I said I'll trade you the book
OK, so you get the book and I get the rest
NO NO NO NO NO NO NO NO, I said I'll give you the book and I'll take the rest
Analysis

Goal-based models negotiate (too) aggressively

- You can have the hat and the book
- I will take the balls and hat
- I need at least 2 balls and the book
- I will take the balls and hat
- Are you reading what I am saying?
- I need 2 balls and the book
- I will take the balls and hat
- No deal
- I will take the balls and hat
Conclusion

Natural Language Negotiations offer **hard** but **important** problem

Planning ahead using **dialogue rollouts** is simple and effective
Any questions?