The Conversational Intelligence Challenge 2

Solution of “Lost in Conversation” team

Sergey Golovanov
Neuromation
sergey_xg@mail.ru

Alexander Tselousov
Independent contributor
aleksander.tselousov@yandex.ru

Speaker: Rauf Kurbanov
Neuromation
JetBrains Research
kurbanov.re@gmail.com

NeurIPS 2018 Competition Track

Challenge

Goal:

A non-goal-oriented dialog system with a persistent persona

Problems:

- Lack of a consistent personality
- Lack of an explicit long-term memory
- Tendency to produce non-specific answers like “I don’t know’’

Ideal solution:

- Simulate a normal conversation
- Learn about the interests of opponent
- Discuss own interests and find common ground

Datasets

PersonaChat (original + revised):

Zhang S. et al. Personalizing Dialogue Agents: I have a dog, do you have pets too?

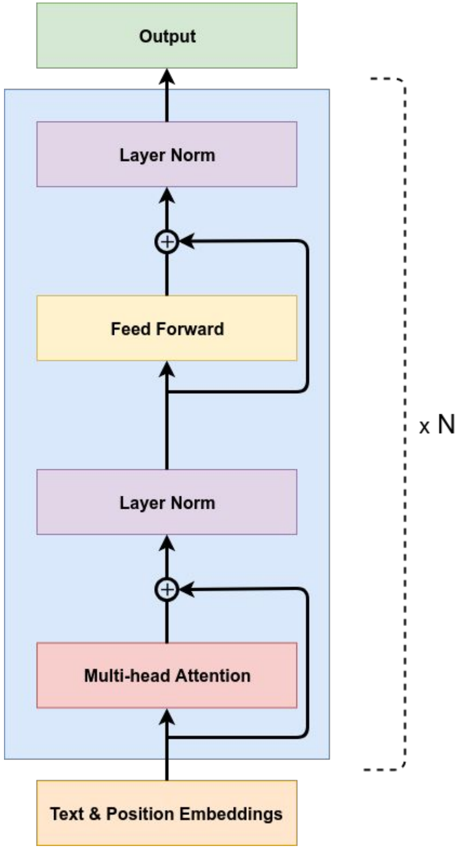
DailyDialog:

Li Y. et al. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset

Reddit comments dataset:

files.pushshift.io/reddit/comments

Base architecture

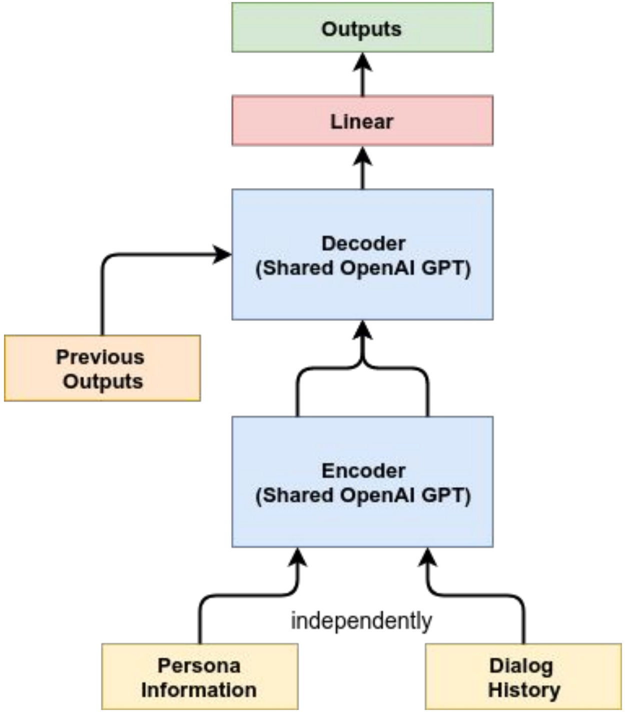


OpenAI GPT[1] :

- BPE vocabulary with 40000 tokens
- Learned position embeddings with 512 positions
- 12 transformer layers
- Multi-head attention with 768 dimensional states and 12 attention heads
- Position-wise feed-forward networks with 3072 dimensional inner states

1. *Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. (2018). Improving language understanding by generative pre-training.*

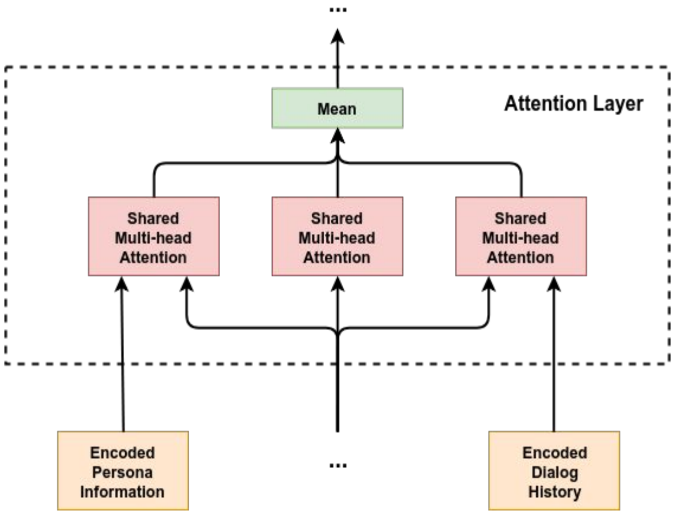
Our architecture



- Shared encoder and decoder - pretrained OpenAI GPT[1]
- Shared pre-softmax linear layer and token embeddings[2]
- Beam-search with length penalty[3] and annealing for improving answer diversity
- Reduction of persona information and dialog history – first and last 512 tokens respectively

1. *Xia Y. et al. Model-level dual learning*
2. *Press O., Wolf L. Using the output embedding to improve language models*
3. *Wu Y. et al. Google's neural machine translation system: Bridging the gap between human and machine translation*

Our architecture: pretrained model


...
Mean
Attention Layer
Shared Multi-head Attention
Shared Multi-head Attention
Shared Multi-head Attention
Encoded Persona Information
...
Encoded Dialog History

Attention layer modifications:

- Shared multi-head attention layers
- Parallel computation of attention for inputs
- Merge of attentions - mean

Learning procedure: loss functions

To train model we used weighted combination of losses[1]:

$$Loss = L_{TokLS} + \lambda_{LM} \cdot L_{LM} + \lambda_{risk} \cdot L_{risk}$$

$$L_{TokLS} = -\sum_i \log P(y_i|y_1, \ldots, y_{i-1}) - D_{KL}(f||P(y_i|y_1, \ldots, y_{i-1}))$$

$$L_{LM} = -\sum_i \log P(y_i|y_1, \ldots, y_{i-1})$$

$$L_{risk} = \sum_{y_{pred} \in B} (1 - f1(y_{target}, y_{pred})) \cdot \frac{p(y_{pred})}{\sum_{y'_{pred} \in B} p(y'_{pred})}$$

First stage:

- $\lambda_{LM} = 0.5$
- $\lambda_{risk} = 0$

- $\lambda_{LM} = 0.1$:
- $\lambda_{risk} = 10$

Beam-search samples

for risk minimization

---

[1] *Edunov S. et al. Classical Structured Prediction Losses for Sequence to Sequence Learning*

Learning procedure: settings

Settings:

- batch size: 256
- learning rate: 6.25e-5
- warmup: 16000
- label smoothing: 0.1
- dropout: 0.1

Training time:

- First stage: ~ week on Nvidia GTX 1080TI
- Second stage: ~ two days on Nvidia GTX 1080TI

Heuristics

Human behavior simulation:

- Sentiment analysis of human replies and adding emoji
- Mistake correction (spell, cycle conversation, ...)
- Questions from retrieval model to find out additional information about an interest

Turned off: decrease automated evaluation metrics

Results: score

	PPL	**Hits@1**	**F1**
Public set	-	0.173	0.179
Private set	-	0.171	0.177

Results: dialog examples

Persona:
I ride the bus to school.
I hate math class.
I am on the soccer team.
My brother is older than me.

H (Human): Hello friend.
B (Bot): Hi how are you today?
H: I am fine, you?
B: I am doing well, just got off work. 😁
H: Where do you work?
B: I'm in school right now. How about you?
H: I am a software developer.
B: That is cool. What do you do for fun? 😄
H: I like to read books.
B: What kind of books do you like?
H: I prefer horrors.
B: Horror movies are the best!
H: I like them too.
B: What is your favorite book?
H: It by Stephen King.
B: I like Stephen King too. Do you work?
H: Right now no.
B: I like to eat pizza with my brother. 😁
H: Bye.
B: Bye.

Persona:
I am a ninja.

H (Human): I like cats.
B (Bot): Cats are my favorite. I have a lot of them. What is your cats name?
H: I call my cat jerk, I hate him.
B: That is funny. Do you live in Japan?
H: No, you?
B: I live in the united states. I moved to America when I was 14. How about you Mars or Jupiter?
H: Mars.
B: Mars is beautiful. But I don't think Mars is as cool as Mars. What are your hobbies?
H: No. Where are you moved to USA from?
B: Japan. It is a great place to live.

Future improvements

What we wanted to do but did not have time and resources for:

- Hybrid of generative and retrieval systems
- Sentence compression
- Attention on sentences
- Dataset augmentation (translate to another language and back)
- Memory and speed optimization
- Critic to evaluate dialog accordance with persona
- …