Machine Learning

Program Assignment 2
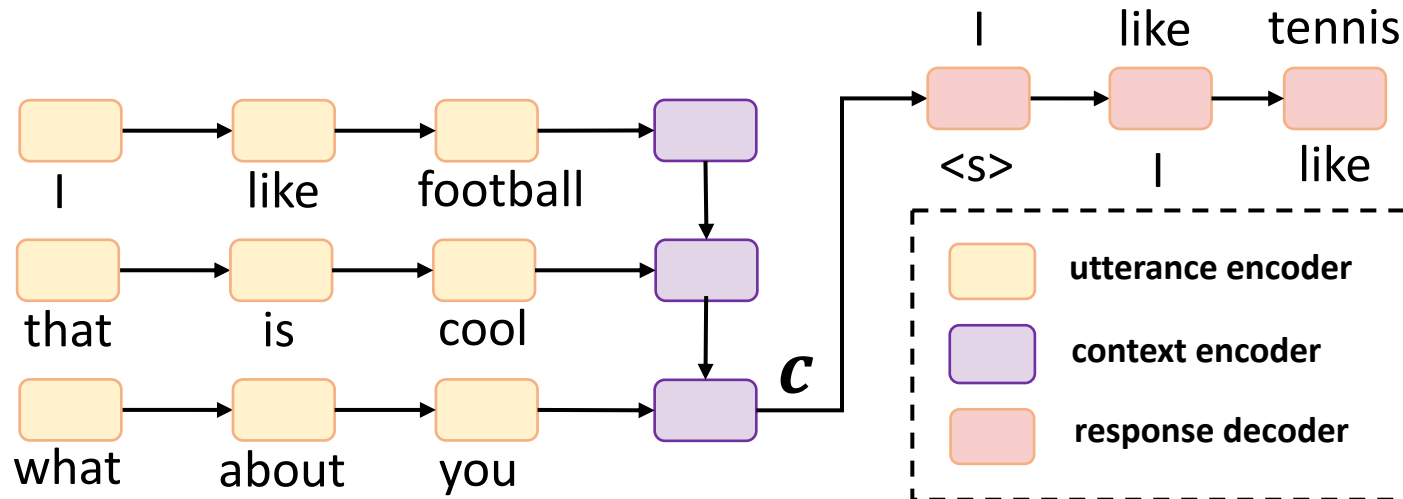
Fall 2020

Instructor: Xiaodong Gu

Goals

- Implement A Chatbot Using Deep Learning

  ✓ achieve high performance

  ✓ learn to implement popular neural language models such as Transformer, etc.

- A presentation to **sell** your product/idea.

- [optional] new ideas for open-domain dialogue generation

Open-Domain Conversations
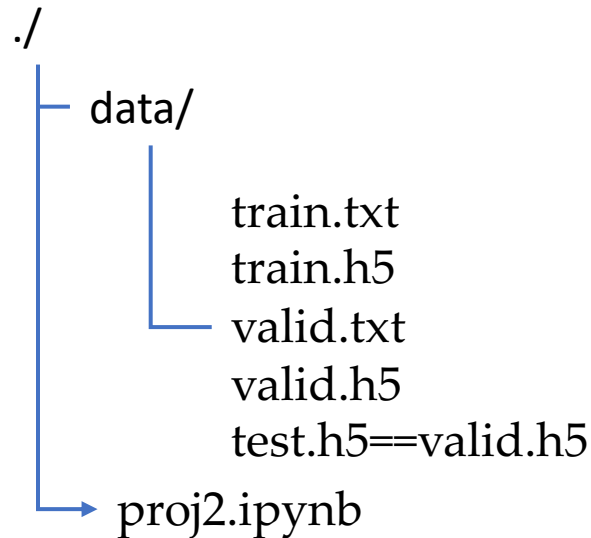
A Sample Architecture

Dataset

We provide a dataset of Chinese chitchat.

Dialog1
1	米芝莲在福州哪儿哈？
2	金山万达
3	好吃咩？

Context

| 4 | 不好吃还不如一般菠萝包跟香港的比真的差多了 |

Next Utterance

| 5 | |

Dialog2
6	第八张是给人家拍的街拍么
7	你好烦
8	

Dialog3
9	你真好意思
10	嫉妒我|
11	

Dialog4
12	好美，哪里
13	乌镇
14	

Dialog5
15	请教，风炉烤月饼温度跟普通烤箱一样嘛？多盘同烤上色如何？谢谢！
16	比普通烤箱上色快我一般不预热直接进去烤降低温度
17	好的，感谢

⋮

Dataset

Download it from Canvas ,

and put the files into a ./data/ folder in your project.

```
./
├── data/
│       │
│       │       train.txt
│       │       train.h5
│       └──     valid.txt
│               valid.h5
│               test.h5==valid.h5
└──→  proj2.ipynb
```

Platform

Python notebook by Tencent's TI-ONE (Recommended)

- Consult in the WeChat group if you have any issues in using the platform.

Python notebook by Anaconda in a local machine

- If you your machine contains a powerful GPU.

A Baseline Implementation

We provide a sample of baseline implementation which is consist of the following sections:

- Libraries
- Utilities
- Configuration
- Data Loader
- Evaluation
- Training Script
- Main function for training
- Main function for test

You can implement your model in the `**Model**` section

You can modify some sections except the 'Evaluation'.

Performance Test

- We provide a test script for you to evaluate your model in a *pseudo* test set (the released test set is just the validation set).
- The TA will verify your results using the same script in a real test set.

Requirements and Key Points

- Data Preprocessing
  - We have provided a data loader for the .h5 file
  - You can adjust the data loader, or design a new one. However, you cannot modify the default test script (the last section of the sample program) as well as the evaluation metrics which the TA will run directly to test your results on a secret test set. So please adapt your code to the test script to avoid mistaken grading.

- Model Implementation

  Modules you can use:

  torch.nn.GRU,. Torch.nn.LSTM, torch.Dataset,

  Modules you **cannot** use:

  torch.nn.Transformer, huggineface's Transformer,…

  A direct invocation of third-party libraries for the whole model will receive a significant penalty.

Presentation

A recorded (录好音的) pptx with the following contents:

Without Idea	With New Ideas
Background	
Related Works	
	Motivation
Technology	Approach
Implementation Details	
Evaluation	
Demo	
task allocation	

Presentation: Background

- No strict requirement. You may consider:
  - Investigate the technical trend of conversation modeling in the industry.
  - Information you have received about chatbot.
  - Your understanding about chatbot.
  - …

Presentation: Related Works

Important technologies (papers) that you have implemented or adopted.

List 2-3 papers and briefly describe the key ideas.
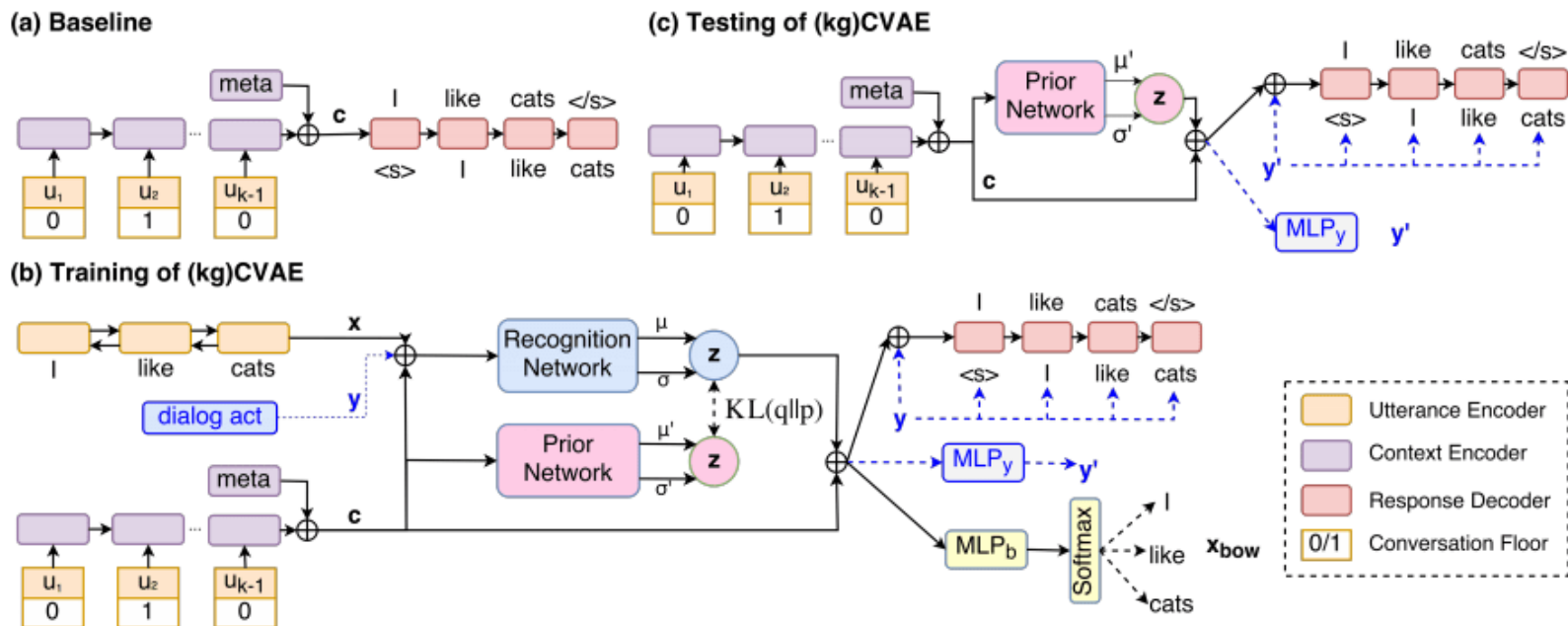
Presentation: Motivation [optional]

If you have an new idea, please show

- what is the main problem of existing approaches?
- how do you address the problem?
- what is the key idea of your model?

Presentation: Approach

Describe your approach using diagrams and descriptions (like how we introduced the Seq2Seq, Attention, Transformer, etc in the class).

For example:

# Presentation: Implementation Details

More details of key components and algorithms (e.g., encoder, decoder, etc) in your implementation.

Presentation: Evaluation

Show the following results.

- curves of loss and bleu scores in the validation set
- comparison of different models (BLEUs) including the provided baseline in the test set.
- brief descriptions about the results and comparisons.
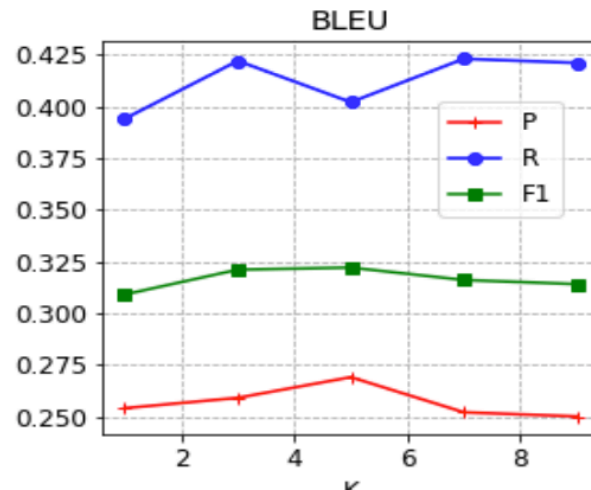
Presentation: Demo

Show some concrete examples (selected context and the generated response) yielded by both yours and the baseline model.

Example

Table 3: Examples of context-response pairs for the neural network models. _eou_ indicates a change of turn. 'Eg.i' means the i-th example.

Context	Examples of Generated Responses	
	CVAE-CO	**DialogWAE-GMP**
thank your for calling world airline. what can I do for you? _eou_	**Eg.**1: i'm afraid i can't find it. **Eg.**2: what's the matter? **Eg.**3: hi, this is my first time.	**Eg.**1: i'd like to make a reservation for you, please **Eg.**2: do you know where i can get to get? **Eg.**3: can you tell me the way to the station?
how much is the rent? _eou_ the rent is $1500 per month.	**Eg.**1: how much is the rent? **Eg.**2: how much is the rent? **Eg.**3: what is the difference?	**Eg.**1: no problem. i'll take it. **Eg.**2: this one is $1.50.50,000 yuan per month. **Eg.**3: that sounds like a good idea.
guess who i saw just now ? _eou_ who? _eou_ john smith. _eou_ that bad egg who took the low road since he was a boy.	**Eg.**1: yes, he is. **Eg.**2: yes, he is **Eg.**3: yes, he is.	**Eg.**1: it is my favorite. **Eg.**2: no, but i didn't think he was able to get married. i had no idea to get her. **Eg.**3: this is not, but it's not that bad. it's just a little bit, but it's not too bad.

Presentation: Task Allocation

Show the contribution of each member.

Grading Scheme

- Performance [10pt]

−top 20%	20~50%	50% -100%	non-sense results
10pt	8pt	6pt	0pt

- Technical soundness [12pt]
  - Easy (e.g., simply attention,): 5pt
  - Medium (e.g., Transformer): 9pt
  - Advanced (e.g., technology newer or beyond Transformer, e.g., diversity objective, GAN, latent variable, RL, etc): 12pt

- Presentation [8pt]

  -

- [optional & bonus] new ideas [+10pt]
  - idea adopted +5pt,
  - and have shown that it works +5pt

Submission

- group project (1-3 members)
- one of the group member submit a single file named 'ID1_NAME1_ID2_NAME2_ID3_NAME3.zip' to Canvas including:
  - code: (proj2.ipynb)
  - trained model (checkpoint_itr0.pkl)
  - presentation (presentation.pptx with voice)

Due date: Dec. 25

Tips

Your programs should be written in such a way that the TA can easily verify the results reported by you.

Your presentation should be clear and comprehensive so that customers (TAs) will buy (give high score to) your product.